ORIGINAL

Machine learning to predict poor school performance in paediatric survivors of intensive care: a population-based cohort study

Patricia Gilholm[1] , Kristen Gibbons[1] , Sarah Brüningk[2,3] , Juliane Klatt[2,3] , Rhema Vaithianathan[4],
Debbie Long[1,5] , Johnny Millar[6,7,8] , Wojtek Tomaszewski[4] and Luregn J. Schlapbach[1,9]* on behalf of the
Australian and New Zealand Intensive Care Society (ANZICS) Centre for Outcomes & Resource Evaluation
(CORE) and ANZICS Paediatric Study Group (ANZICS PSG)

Abstract

Purpose: Whilst survival in paediatric critical care has improved, clinicians lack tools capable of predicting long-term outcomes. We developed a machine learning model to predict poor school outcomes in children surviving intensive care unit (ICU).

Methods: Population-based study of children < 16 years requiring ICU admission in Queensland, Australia, between 1997 and 2019. Failure to meet the National Minimum Standard (NMS) in the National Assessment Program-Literacy and Numeracy (NAPLAN) assessment during primary and secondary school was the primary outcome. Routine ICU information was used to train machine learning classifiers. Models were trained, validated and tested using stratified nested cross-validation.

Results: 13,957 childhood ICU survivors with 37,200 corresponding NAPLAN tests after a median follow-up duration of 6 years were included. 14.7%, 17%, 15.6% and 16.6% failed to meet NMS in school grades 3, 5, 7 and 9. The model demonstrated an Area Under the Receiver Operating Characteristic curve (AUROC) of 0.8 (standard deviation SD, 0.01), with 51% specificity to reach 85% sensitivity [relative Area Under the Precision Recall Curve (rel-AUPRC) 3.42, SD 0.06]. Socio-economic status, illness severity, and neurological, congenital, and genetic disorders contributed most to the predictions. In children with no comorbidities admitted between 2009 and 2019, the model achieved a AUROC of 0.77 (SD 0.03) and a rel-AUPRC of 3.31 (SD 0.42).

Conclusions: A machine learning model using data available at time of ICU discharge predicted failure to meet minimum educational requirements at school age. Implementation of this prediction tool could assist in prioritizing patients for follow-up and targeting of rehabilitative measures.

Keywords: Child, Intensive care, Machine learning, Neurodevelopment, School

*Correspondence: luregn.schlapbach@kispi.uzh.ch
[9] Department of Intensive Care and Neonatology, and Children's Research
Center, University Children's Hospital Zurich, Steinwiesstrasse 75,
8032 Zurich, Switzerland
Full author information is available at the end of the article

Introduction

Progress in the field of paediatric intensive care over the past decades has led to a reduction of in-hospital mortality to as little as 2.5% even for complex conditions such as congenital heart disease or cancer [1–3]. However, critical illness during childhood occurs at a vulnerable period

🍃 Springer

of brain development, and neurological injury may result from disease, complications or treatment-related mechanisms, for example inadequate cerebral oxygen supply during shock or drug-related toxicity [4–6]. Families of critically ill children, clinicians, and researchers consider survival with good long-term neurodevelopment as a priority for care, benchmarking, and research [7]. The ability of a child to meet minimum requirements in primary or secondary school represents a desirable outcome from the family, healthcare provider and societal perspectives and translates into a high chance to ultimately learn a profession, earn an income and lead an independent life in adulthood. Yet, most paediatric intensive care unit (PICU) survivors are not offered follow-up beyond hospital discharge due to lack of long-term follow-up resources. Currently, there are no models available enabling the prediction of long-term neurodevelopmental outcomes which permit risk stratification to target post-discharge rehabilitation measures for children most likely to benefit. Machine learning approaches to make unbiased use of large datasets carry great promise to improve prediction of complex outcomes in heterogeneous populations [8, 9].

We hypothesised that routinely collected healthcare data available at time of discharge from the intensive care unit (ICU) could train machine learning algorithms to predict which children are likely to fail to meet minimum requirements at school age. Through data-driven feature selection incorporated within the machine learning algorithms, we also aimed to identify which variables were most predictive of poor educational outcomes.

Methods

Study design and overview

This is a state-wide data linkage study incorporating data from the Australian and New Zealand Paediatric Intensive Care Registry (ANZPICR) [10] and the National Assessment Program-Literacy and Numeracy (NAPLAN; https://www.nap.edu.au/). Ethical approval was obtained from the Children's Health Queensland Human Research Ethics Committee (HREC/16/QRCH/255). Reporting follows the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) guidelines [11].

Study population

Eligible children were aged below 16 years when admitted to an ICU in the state of Queensland, Australia, between 1st January 1997 and 31st December 2019 and had available NAPLAN educational outcome data. Children who died before NAPLAN testing and those where the NAPLAN testing occurred before ICU admission were excluded.

Take-home message

In this population-based cohort study including 13,957 children admitted to the intensive care unit (ICU), a machine learning model using data available at time of ICU discharge predicted failure to meet minimum requirements at primary and secondary school age with a median follow-up duration of 6 years. The proposed model provides a tool for healthcare providers to stratify paediatric ICU patients in relation to their risk of long-term sequelae, which can assist the design of early rehabilitative measures and inform the prioritisation of patients for follow-up

Data sources

Data were extracted from ANZPICR which collects data from all PICU admissions in Australia, including demographics, admission details, diagnoses, procedures, physiology and specific ICU therapies. Contribution to ANZPICR is a mandatory requirement, and > 90% of all PICU admissions have been captured in the registry. Through linkage with the state-wide death registry, children who were still alive by their first projected NAPLAN test date were identified.

NAPLAN was established in 2008 and is a nationwide annual assessment of students' literacy and numeracy skills, undertaken in primary (Grades 3 and 5) and secondary school (Grades 7 and 9). Children sit their NAPLAN tests usually at age of 8.5 years in Grade 3 and 14.5 years in Grade 9 [12]. Australia has a mixed school system, and NAPLAN is assessed across public and private schools. NAPLAN contains five domains: Grammar and Punctuation, Spelling, Writing, Reading and Numeracy. Performance on the Reading and Numeracy domains is considered most relevant [13]. For every grade, thresholds for National Minimum Standards (NMS) are defined by the Australian Curriculum Assessment and Reporting Authority which capture the minimum acceptable standard of educational achievement without which a student is expected to fail to make sufficient progress at school. Children are exempt from NAPLAN testing if they have a significant neurocognitive or neuromotor disability or other coexisting conditions that severely limit their capacity to participate in the tests to the extent that they are expected to fail. Exempt children are by rule classified as failing to meet the NMS. Children are classified as Absent or Withdrawn if they do not complete the test as a result of being absent from school on the test day or withdrawn by their parents/carers.

Data preparation

The ANZPICR dataset was linked to the individual NAPLAN records provided by the Queensland Curriculum and Assessment Authority using deterministic and probabilistic linkage performed by Queensland Health's Statistical Services Branch [14]. Description of

the selection of the analysed cohort and pre-processing of the data (including data transformations, missing data imputation and the construction of principal component scores) is described in Supplementary Methods. Children who were marked as Absent or Withdrawn from both the Reading and Numeracy domains, and duplicate NAPLAN tests, were excluded. For children with multiple ICU admissions, the child's last ICU admission prior to the corresponding NAPLAN test was used for analysis and sensitivity analyses when using the first admission and the worst admission (based on the highest Paediatric Index of Mortality 2 [PIM2] score) were performed.

Outcomes

The primary outcome was failing to meet the NMS on both the Reading and Numeracy domains for the corresponding grade [14]. Secondary outcomes were failing to meet the NMS on the Reading or the Numeracy domains, failing to meet the NMS on any of the five domains, failing to meet the NMS on at least two domains, and failing to meet the NMS on all domains.

Predictors

We selected all predictors that were available using routine ANZPICR data at time of ICU discharge encompassing demographics, admission information, physiology, specific therapies and diagnostic/procedural codes (Supplementary Methods). The ANZPICR contains over 500 diagnostic and procedural codes used to record the principal, underlying and associated diagnoses for each admission. Dimensionality reduction was performed using principal component analysis (Supplementary Table 1). The principal component (PC) scores were extracted and used as predictors in place of the individual codes. Socio-economic status (SES) was assessed using both the Socio-Economic Indexes for Areas-Index of Relative Socio-economic Disadvantage (IRSD) deciles derived from the postcode recorded at time of admission as well as an SES category variable calculated from parental occupation and education variables (Supplementary Methods) [14–17].

Machine learning models

The following machine learning models commonly used for classification tasks were compared: Logistic Regression with L1 regularisation, [18] Random Forest [19], Gradient Boosting Machine[20] and a Multi-Layer Perceptron neural network [21]. All modelling was performed in Python (version 3.8.8) using the scikit-learn and LightGBM packages and the code for each model is available online (https://github.com/TrishGilholm/ML_prediction_school_outcomes).

In order to control bias and avoid overfitting, the models were trained, validated and tested using five-fold stratified nested cross-validation (Supplementary Methods). We evaluated the performance of the four models within each Grade, separately. The performance of each model was summarized by calculating the mean and standard deviation of the Area Under the Receiver Operating characteristic Curve (AUROC) and Area Under the Precision Recall Curve (AUPRC) across the five test folds. The mean and the standard deviation for the relative AUPRC (the AUPRC divided by the prevalence of the outcome) were also calculated to enable comparison of the AUPRCs across grades, where the prevalence of the outcomes differs. The best performing model out of the four models compared was a priori defined as the model that had the highest mean relative AUPRC when averaged across the four grades. Secondary outcomes were assessed for each grade using the best performing model only.

A final model, combining all four grades, was derived. In order to keep the observations independent, the data comprising the next available NAPLAN test result closest to the last recorded ICU admission for each child was used. The predictive performance of the combined model was evaluated across a range of clinical subgroups. The according sensitivity, specificity, positive likelihood ratio and negative likelihood ratio were also calculated at thresholds corresponding to 75%, 80%, 85% and 90% sensitivity based on published recommendations[22, 23]. Sensitivity analyses were conducted using 33 ICU-related modifiable factors as predictors only (demographic variables and diagnostic and procedural PC scores removed) and on a modified composite outcome of death after ICU or failure to meet NMS. SHapley Additive exPlanation (SHAP) plots [24] were produced for all models to identify the top 20 predictors which contributed most to the model predictions. The normalized importance values were aggregated across the five test folds by calculating the mean and the standard deviation of the importance values for each variable.

Results

Patients and outcomes

During the study period, a total of 28,882 children were admitted to ICU in Queensland. Of these, 18,190 were of the required age to be expected to participate in NAPLAN testing at least once. For 4233 (23.3%) of these children, no NAPLAN data were available (Supplementary Methods, Fig. 1). In total, 13,957 children representing 21,304 ICU admissions had corresponding NAPLAN data encompassing 37,200 NAPLAN tests. Children with linked NAPLAN data were older, more likely to be an elective admission and had lower predicted mortality in
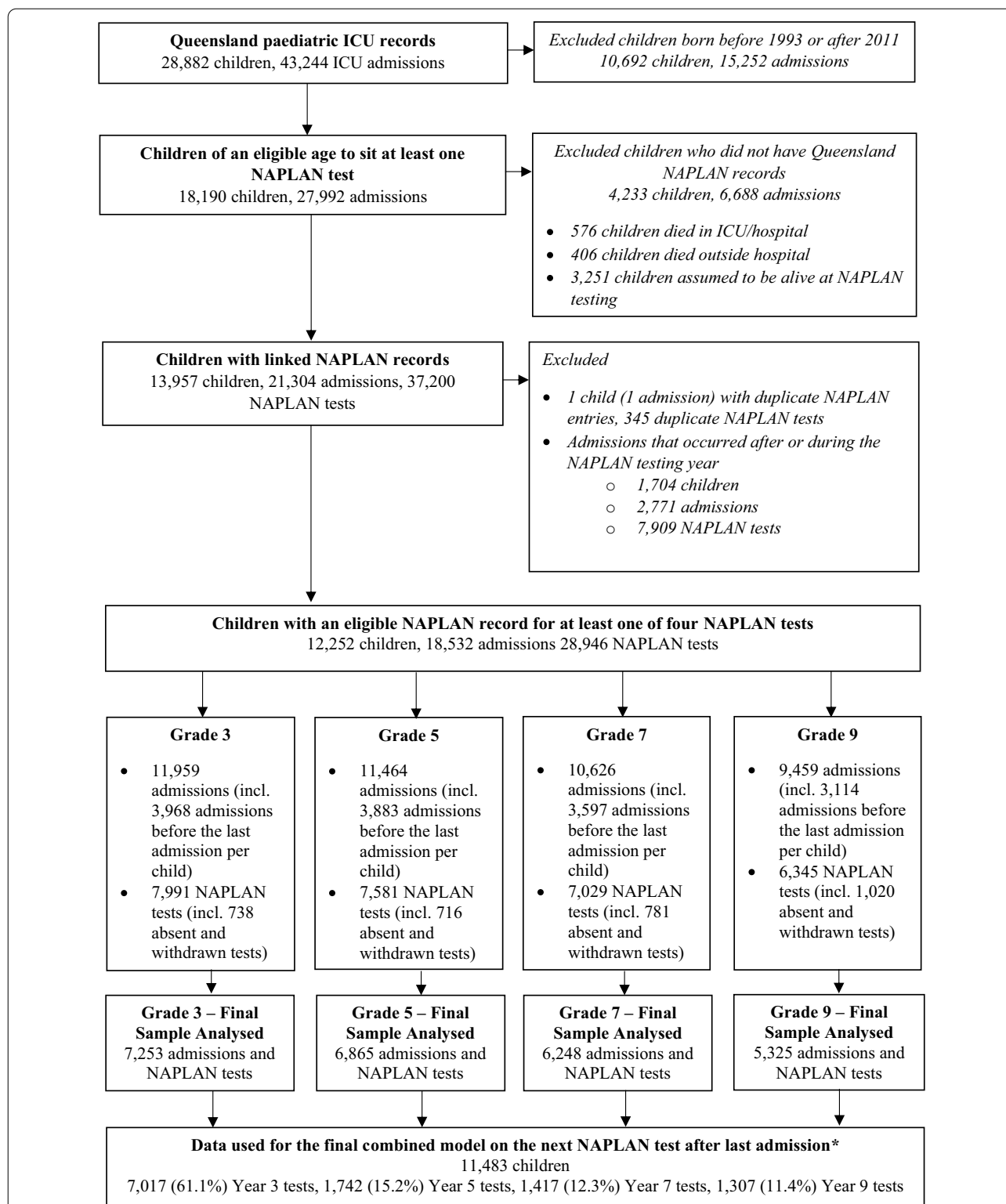
Queensland paediatric ICU records
28,882 children, 43,244 ICU admissions

Excluded children born before 1993 or after 2011
10,692 children, 15,252 admissions

Children of an eligible age to sit at least one NAPLAN test
18,190 children, 27,992 admissions

Excluded children who did not have Queensland NAPLAN records
4,233 children, 6,688 admissions

- *576 children died in ICU/hospital*
- *406 children died outside hospital*
- *3,251 children assumed to be alive at NAPLAN testing*

Children with linked NAPLAN records
13,957 children, 21,304 admissions, 37,200 NAPLAN tests

Excluded

- *1 child (1 admission) with duplicate NAPLAN entries, 345 duplicate NAPLAN tests*
- *Admissions that occurred after or during the NAPLAN testing year*
  - *1,704 children*
  - *2,771 admissions*
  - *7,909 NAPLAN tests*

Children with an eligible NAPLAN record for at least one of four NAPLAN tests
12,252 children, 18,532 admissions 28,946 NAPLAN tests

Grade 3
- 11,959 admissions (incl. 3,968 admissions before the last admission per child)
- 7,991 NAPLAN tests (incl. 738 absent and withdrawn tests)

Grade 5
- 11,464 admissions (incl. 3,883 admissions before the last admission per child)
- 7,581 NAPLAN tests (incl. 716 absent and withdrawn tests)

Grade 7
- 10,626 admissions (incl. 3,597 admissions before the last admission per child)
- 7,029 NAPLAN tests (incl. 781 absent and withdrawn tests)

Grade 9
- 9,459 admissions (incl. 3,114 admissions before the last admission per child)
- 6,345 NAPLAN tests (incl. 1,020 absent and withdrawn tests)

Grade 3 – Final Sample Analysed
7,253 admissions and NAPLAN tests

Grade 5 – Final Sample Analysed
6,865 admissions and NAPLAN tests

Grade 7 – Final Sample Analysed
6,248 admissions and NAPLAN tests

Grade 9 – Final Sample Analysed
5,325 admissions and NAPLAN tests

Data used for the final combined model on the next NAPLAN test after last admission*
11,483 children
7,017 (61.1%) Year 3 tests, 1,742 (15.2%) Year 5 tests, 1,417 (12.3%) Year 7 tests, 1,307 (11.4%) Year 9 tests

Fig. 1 Participant flow diagram demonstrating the selection of the analyzed sample for each of the four school grades and numbers included the final combined model. *ANZPICR* Australia and New Zealand Paediatric Intensive Care Registry, *NAPLAN* National Assessment Program-Literacy and Numeracy, *ICU* intensive care unit; *For the combined model, only one ICU admission and one NAPLAN testing year outcome were used per child. Specifically, the NAPLAN test following the last recorded and eligible ICU admission per child was used for the combined model. As most children are admitted to ICU before they are eight years old, the NAPLAN test included most commonly in the final model was the Grade 3 test.

comparison to those with no linked NAPLAN records (Supplementary Tables 2 and 3). After exclusions due to absent and withdrawn students (Supplementary Table 4), the final analyzed sample sizes were 7253 children for Grade 3, 6865 children for Grade 5, 6248 children for Grade 7 and 5325 children for Grade 9, respectively (Table 1). 14.7%, 17%, 15.6% and 16.6% of children failed to meet the NMS on both the reading and numeracy domains in Grade 3, 5, 7 and 9, respectively, which was several fold greater than the overall Queensland population (Supplementary Table 5, Supplementary Fig. 1).

Model development

When comparing the performance of the four machine learning prediction models across the grades, the mean AUROCs ranged from 0.76 (SD 0.01) to 0.80 (SD 0.02; Supplementary Fig. 2; Supplementary Table 6). The mean relative AUPRCs ranged from 3.0 (SD 0.11) to 3.67 (SD 0.12). Sensitivity analyses using the first and worst ICU admissions resulted in similar performance (Supplementary Fig. 3). The Gradient Boosting Machine was selected as the best performing model, with similar performance achieved across grades (Table 2, Supplementary Figs. 4 and 5, Supplementary Table 6). Variables relating to demographics (SES, IRSD decile, age at admission), illness severity (number of ICU admissions, ICU length of stay, base excess, systolic blood pressure and partial O_2 pressure), and diagnostic and procedural code PC clusters relating to neurological, congenital, and genetic disorders contributed most to the predictions (Supplementary Fig. 6). High SES was identified as a consistent protective factor. When assessing secondary outcomes, the performance of the model increased with a higher number of failed NAPLAN domains, with the best predictions produced for failing to meet NMS on all five domains (Supplementary Tables 7–10).

Final combined model

The final combined model comprised 11,483 admissions, of which 61.1%, 15.2%, 12.3% and 11.4% were of NAPLAN Grades 3, 5, 7 and 9, respectively, aligning with the age distribution of children requiring ICU admissions. The median time between ICU discharge and NAPLAN test was 6 years (IQR 2, 8) and 16.7% failed to meet the NMS on the reading and numeracy domains. The combined model demonstrated good predictive performance (mean AUROC 0.80 [SD 0.01], mean relative AUPRC 3.42 [0.07]) (Fig. 2A, B). At 85% sensitivity, the model reached a specificity of 51% (SD 3%), a positive likelihood ratio of 1.75 (SD 0.14), and a negative likelihood ratio of 0.29 (SD 0.02) (Table 2). In subgroup analyses, the model resulted in mean AUROCs of >0.70 for all subgroups except prematurity, and demonstrated

the best predictive performance for pneumonia, chronic neurological conditions, admission years between 2009 and 2019, respiratory infections, and children with any comorbidities (Fig. 2C, Supplementary Table 11). In children with no comorbidities admitted between 2009 and 2019, the model achieved a mean AUROC of 0.77 (SD 0.03), and a mean relative AUPRC of 3.31 (SD 0.42).

Risk factors with the highest mean normalized importance included demographics, illness severity, and diagnostic and procedural code PC clusters associated with neurological, infectious, respiratory, congenital, and genetic disorders (Fig. 3). A sensitivity analysis using ICU-related factors only as predictors demonstrated reduced predictive performance (mean AUROC 0.67 [SD 0.02], mean relative AUPRC 1.91 [SD 0.11], Supplementary Fig. 7), but highlighted potentially modifiable factors contributing most to the outcome, most of which were also identified in the main model (e.g. higher number of previous ICU admissions, longer total ICU length of stay and lower systolic blood pressure on admission). Sensitivity analyses using the composite outcome death after PICU (565/12,029 [4.7%] deaths) or failure to meet NMS resulted in improved predictive performance overall and in most subgroups, and revealed similar risk factors (mean AUROC 0.84 [SD 0.01], mean relative AUPRC 3.65 [SD 0.06]), Supplementary Figs. 8 and 9).

Discussion

In this population-based study of a large cohort of children admitted to ICU, a machine learning model using data available at time of ICU discharge predicted failure to meet minimum requirements at primary and secondary school age, at a median follow-up of 6 years. The model built for all survivors of PICU was capable of predicting failure to meet minimum educational requirements even in children without comorbidities, as well as in children with diverse conditions such as pneumonia or trauma, with best performance observed in patients admitted to ICU between 2009 and 2019.

Contrary to a multitude of models predicting early mortality in critically ill children [10, 25], clinicians currently lack tools capable of predicting long-term outcomes in children surviving ICU. Recent prediction models on non-mortality outcomes focus on health-related quality of life measured through standardized parent-reported questionnaires obtained within twelve months of ICU discharge and were primarily built on specific research cohorts not necessarily generalisable to broader populations [26–28]. Measures of quality of life alone may not adequately capture cognitive and functional performance, which is essential for educational and ultimately, professional achievement [29–32]. Comprehensive neuropsychological face-to-face testing

Table 1 Demographic and clinical characteristics of PICU survivors assessed within each NAPLAN school grade cohort

Characteristic	Grade 3 N = 7253	Grade 5 N = 6865	Grade 7 N = 6248	Grade 9 N = 5325
Demographics				
Gender (male)	4227 (58.3%)	3955 (57.6%)	3584 (57.4%)	3040 (57.1%)
Aboriginal and Torres Strait Islander	304 (4.2%)	209 (3%)	162 (2.6%)	132 (2.5%)
Age at admission (months)	19 (4, 52)	26 (6, 67)	38 (8, 92)	61 (14, 125)
Index of Relative Socio-economic Disadvantage (IRSD) Decile	6 (3, 8)	6 (3, 8)	6 (3, 8)	6 (3, 8)
Low SES	1904 (26.3%)	1804 (26.3%)	1564 (25%)	1195 (22.4%)
Medium SES	2228 (30.7%)	2051 (29.9%)	1876 (30%)	1568 (29.4%)
High SES	2217 (30.6%)	2090 (30.4%)	1927 (30.8%)	1731 (32.5%)
No information	904 (12.5%)	920 (13.4%)	881 (14.1%)	831 (15.6%)
Time between ICU discharge and NAPLAN test (years)	7 (4, 8)	8 (5, 10)	9 (5, 12)	9 (4, 13)
Comorbidities				
No comorbidities	3607 (49.7%)	3558 (51.8%)	3362 (53.8%)	2967 (55.7%)
Prematurity	351 (4.8%)	226 (3.3%)	99 (1.6%)	43 (0.8%)
Congenital heart disease	1265 (17.4%)	1157 (16.9%)	1019 (16.3%)	785 (14.7%)
Congenital syndrome	449 (6.2%)	404 (5.9%)	373 (6%)	317 (6%)
Oncologic conditions	182 (2.5%)	171 (2.5%)	162 (2.6%)	157 (2.9%)
Chronic respiratory conditions	823 (11.3%)	672 (9.8%)	550 (8.8%)	379 (7.1%)
Chronic neurological conditions	482 (6.6%)	514 (7.5%)	514 (8.2%)	501 (9.4%)
Infections				
Respiratory infections	1312 (18.1%)	1016 (14.8%)	764 (12.2%)	545 (10.2%)
Pneumonia or Pneumonitis	441 (6.1%)	381 (5.5%)	295 (4.7%)	227 (4.3%)
Bronchiolitis	650 (9%)	460 (6.7%)	310 (5%)	194 (3.6%)
Invasive infection	804 (11.1%)	720 (10.5%)	581 (9.3%)	484 (9.1%)
Sepsis/Septic shock	305 (4.2%)	269 (3.9%)	209 (3.3%)	182 (3.4%)
Other				
Asthma	417 (5.7%)	396 (5.8%)	363 (5.8%)	304 (5.7%)
Diabetic ketoacidosis	146 (2%)	191 (2.8%)	240 (3.8%)	281 (5.3%)
Trauma				
All trauma	773 (10.7%)	865 (12.6%)	901 (14.4%)	860 (16.2%)
Head Trauma	299 (4.1%)	351 (5.1%)	398 (6.4%)	373 (7%)
Elective surgery				
ENT surgery	809 (11.2%)	722 (10.5%)	602 (9.6%)	406 (7.6%)
Congenital Heart Surgery	1182 (16.3%)	1192 (17.4%)	1011 (16.2%)	858 (16.1%)
Illness severity and treatment				
ICU length of stay—last admission (hours)	24 (17, 50)	23 (17, 48)	23 (16, 46)	22 (16, 45)
Invasive ventilation	2877 (39.7%)	2833 (41.3%)	2575 (41.2%)	2152 (40.4%)
Time intubated—last admission (hours)	0 (0, 16)	0 (0, 16)	0 (0, 15)	0 (0, 14)
Paediatric Index of Mortality (PIM)-2 (%)	0.5 (0.1, 1.2)	0.5 (0.1, 1.2)	0.6 (0.1, 1.3)	0.6 (0.1, 1.3)
Inter-hospital transfer preceding ICU admission	1538 (21.2%)	1417 (20.6%)	1264 (20.2%)	1086 (20.4%)
Elective admission	3249 (44.8%)	3118 (45.4%)	2822 (45.2%)	2364 (44.4%)
Number of previous ICU admissions				
0	5645 (77.8%)	5327 (77.6%)	4851 (77.6%)	4149 (77.9%)
1	914 (12.6%)	858 (12.5%)	773 (12.4%)	642 (12.1%)
2–4	555 (7.7%)	532 (7.7%)	499 (8%)	428 (8%)
5 or more	139 (1.9%)	148 (2.2%)	125 (2%)	106 (2%)

Data presented correspond to the last ICU admission before the corresponding NAPLAN test per child. Counts and percentages, and median and interquartile ranges are provided

Children can have up to four NAPLAN tests so the same child can be represented in multiple grades. However, the last PICU admission before the corresponding NAPLAN test may be different for children who have had PICU admissions between NAPLAN test years

ICU Intensive Care Unit, *NAPLAN* National Assessment Program-Literacy and Numeracy, *SES* socioeconomic status, *ENT* ears, nose and throat

Table 2 Specificity, positive likelihood ratio and negative likelihood ratio, corresponding to the predefined sensitivity thresholds of 75%, 80%, 85% and 90% for each of the individual grade models and the combined model

Sensitivity threshold	Mean specificity % (SD)	Mean PLR (SD)	Mean NLR (SD)
Grade 3			
75%	65% (4%)	2.16 (0.21)	0.38 (0.02)
80%	57% (4%)	1.89 (0.21)	0.35 (0.03)
85%	49% (2%)	1.65 (0.1)	0.3 (0.01)
90%	42% (4%)	1.56 (0.1)	0.24 (0.02)
Grade 5			
75%	65% (3%)	2.12 (0.14)	0.39 (0.01)
80%	58% (3%)	1.94 (0.12)	0.34 (0.01)
85%	52% (3%)	1.78 (0.11)	0.28 (0.01)
90%	42% (3%)	1.55 (0.07)	0.24 (0.01)
Grade 7			
75%	67% (5%)	2.29 (0.35)	0.37 (0.03)
80%	58% (5%)	1.95 (0.29)	0.34 (0.04)
85%	49% (3%)	1.68 (0.11)	0.31 (0.02)
90%	37% (4%)	1.24 (0.49)	0.26 (0.03)
Grade 9			
75%	64% (6%)	2.16 (0.33)	0.39 (0.03)
80%	57% (5%)	1.87 (0.24)	0.36 (0.03)
85%	45% (8%)	1.58 (0.23)	0.34 (0.06)
90%	33% (8%)	1.36 (0.19)	0.31 (0.06)
Combined model			
75%	68% (2%)	2.35 (0.16)	0.37 (0.01)
80%	60% (2%)	2.03 (0.12)	0.33 (0.01)
85%	51% (3%)	1.75 (0.14)	0.29 (0.02)
90%	40% (3%)	1.5 (0.09)	0.25 (0.02)

The mean and the standard deviation across cross-validation folds are reported for each statistics

PLR positive likelihood ratio, *NLR* negative likelihood ratio, *SD* standard deviation

represents the gold standard to assess long-term outcomes in children, but remains highly resource-intensive and is often not available outside highly selected research cohorts [33–35]. Verlinden et al. followed up 614 children on average 1.9 years post-PICU and observed inferior scores across several neuropsychological outcomes compared to healthy children [34]. Our study including educational encounters of 13,957 PICU survivors provides unprecedented power coupled with the strengths of a population-based design across a prospective mandatory ICU registry known for its high internal and external validity [36]. The overall incidence of 16.7% failure to meet NMS corroborates a previous Finnish study, where 13% of 753 PICU survivors were reported to have major difficulties in school, using the Strengths and Difficulties questionnaire [37].

Prediction modelling was restricted to routine healthcare data available at time of ICU discharge and demonstrated comparable results across the four school grades, reaching a specificity of 51% at 85% sensitivity. The American Academy of Pediatrics Committee on Children with Disabilities stated requirements for neurodevelopmental screening tools [22, 23]; however, these are predominantly applied to tests identifying disability at time of testing, rather than predicting educational outcomes several years later. The final model revealed three key groups of variables contributing to poor educational outcomes: sociodemographic information, disease severity and disease clusters. The strong role of SES is supported by a broad body of literature across preterm and paediatric patients [14, 17, 38]. This finding demonstrates the particular vulnerability of children from lower socioeconomic backgrounds and highlights the importance to target post-discharge measures to such disadvantaged populations [39–42]. Several severity measures emerged as key predictors, including number and duration of ICU admissions, low systolic blood pressure, base excess and cardiorespiratory arrest. We posit that these variables serve as proxy markers of brain damage associated with severe illness, risk of hypotension or hypoxia during ICU, as well as prolonged exposure to side effects from ICU therapies such as sedation [6, 43]. Some of these variables are potentially modifiable.

The heterogeneity of ICU diseases previously posed obstacles for accurate non-mortality prediction models. Using principal components analysis on diagnostic and procedural codes, we were able to define combinations with particular impact on long-term outcomes. As variations of the ANZPICR are used in Oceania, Europe and South America, implementation of such a model may permit individualized stratification of patients at time of ICU discharge in relation to their risk of long-term sequelae. However, the generalizability of our findings is limited by the Australian educational outcomes assessed in our study, the fact that the cohort spans across 20 years of ICU care [1], and considering the lack of NAPLAN data in 23% of survivors due to interstate movement, home schooling and linkage failures. Of note, the predictive performance improved substantially when assessing children admitted in the more recent decade, implying that prospective calibration of the model may further enhance its accuracy. Sensitivity analyses predicting death after ICU or failure to meet NMS improved the diagnostic performance, demonstrating the robustness of the model. Whilst composite outcomes of death or disability have been used in the ICU literature [44], the earlier timing of death, fundamentally distinct implications for families, and differences in actionable measures (such as rehabilitation) in our view favour the design of tools
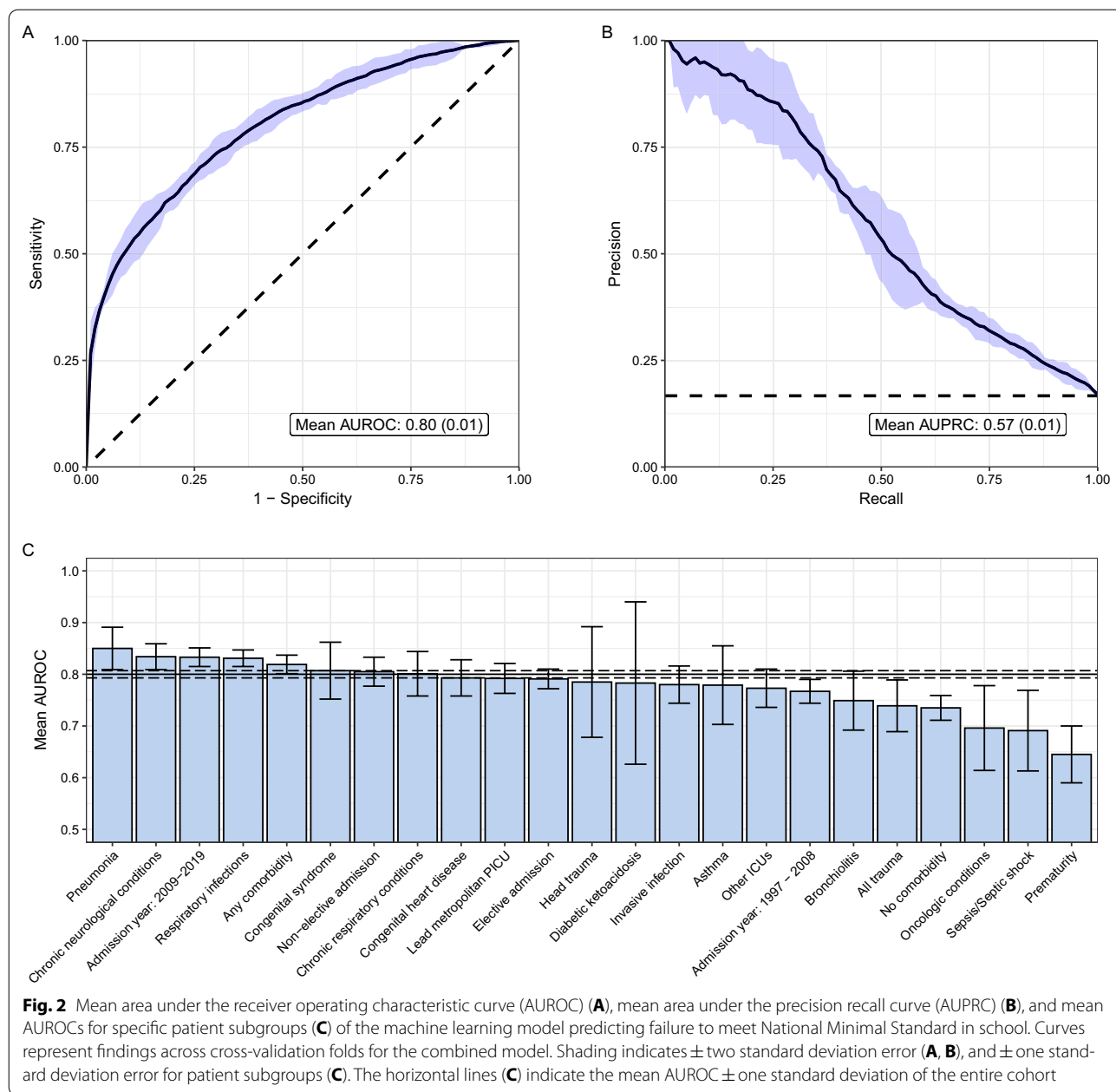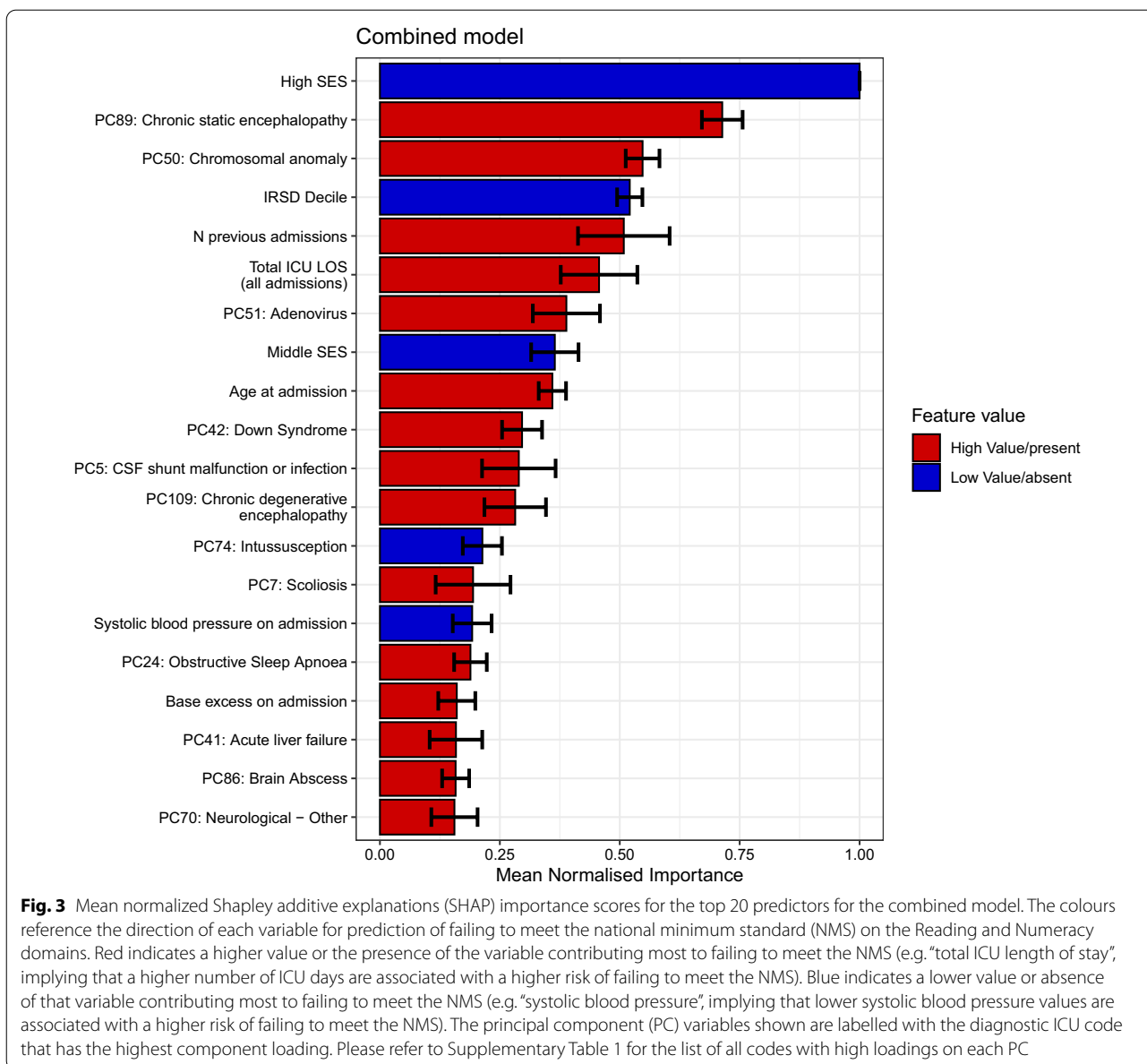
Fig. 2 Mean area under the receiver operating characteristic curve (AUROC) (**A**), mean area under the precision recall curve (AUPRC) (**B**), and mean AUROCs for specific patient subgroups (**C**) of the machine learning model predicting failure to meet National Minimal Standard in school. Curves represent findings across cross-validation folds for the combined model. Shading indicates ± two standard deviation error (**A**, **B**), and ± one standard deviation error for patient subgroups (**C**). The horizontal lines (**C**) indicate the mean AUROC ± one standard deviation of the entire cohort

predicting failure to meet NMS. Whether implementing the algorithm will allow to select patients for more resource-intensive follow-up or facilitate the enrichment of patients more likely to benefit from targeted rehabilitative or educational interventions [45, 46] needs to be tested by future studies.

Several additional limitations need to be considered. Due to the unique availability of the included educational databases, external validation was not feasible. Instead, we used nested cross-validation to avoid overfitting. Although NAPLAN scores correlate with standardized

intelligence quotient measures, as well as with school assessments undertaken in the U.S. and in Singapore [47–49], they are not internationally used limiting direct comparisons with other school systems. Inherent to the nature of PICU admissions, the vast majority occurred before school entry, thus prohibiting before-after analyses. Finally, preterm infants which are usually admitted to neonatal rather than PICUs represented only 3% of the cohort and unique developmental risks related to prematurity were not captured by the ANZPICR.

Fig. 3 Mean normalized Shapley additive explanations (SHAP) importance scores for the top 20 predictors for the combined model. The colours reference the direction of each variable for prediction of failing to meet the national minimum standard (NMS) on the Reading and Numeracy domains. Red indicates a higher value or the presence of the variable contributing most to failing to meet the NMS (e.g. "total ICU length of stay", implying that a higher number of ICU days are associated with a higher risk of failing to meet the NMS). Blue indicates a lower value or absence of that variable contributing most to failing to meet the NMS (e.g. "systolic blood pressure", implying that lower systolic blood pressure values are associated with a higher risk of failing to meet the NMS). The principal component (PC) variables shown are labelled with the diagnostic ICU code that has the highest component loading. Please refer to Supplementary Table 1 for the list of all codes with high loadings on each PC

In conclusion, a machine learning algorithm using data available at time of ICU discharge provides a tool capable of predicting which children are likely to fail meeting minimum educational requirements during primary and secondary school. Prospective evaluation of this prediction tool can assist the design of future implementation strategies which may enable personalized prioritization for follow-up and early rehabilitative measures.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1007/s00134-023-07137-1.

Author details
[1] Child Health Research Centre, The University of Queensland, Brisbane, QLD, Australia. [2] Department of Biosystems Science and Engineering, ETH Zurich, 4058 Basel, Switzerland. [3] SIB Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland. [4] Institute for Social Science Research, The University of Queensland, Brisbane, QLD, Australia. [5] School of Nursing, Centre for Healthcare Transformation, Queensland University of Technology, Brisbane, QLD, Australia. [6] Paediatric Intensive Care Unit, The Royal Children's Hospital, Melbourne, VIC, Australia. [7] The Australian and New Zealand Intensive Care Society (ANZICS) Centre for Outcome and Resource Evaluation (CORE), ANZICS House, Melbourne, VIC, Australia. [8] Murdoch Children's Research Institute, Melbourne, VIC, Australia. [9] Department of Intensive Care and Neonatology, and Children's Research Center, University Children's Hospital Zurich, Steinwiesstrasse 75, 8032 Zurich, Switzerland.

Sainath Raman, Debbie Long, Luregn Schlapbach, Kristen Gibbons, Shane George, Puneet Singh, Vicky Smith, Warwick Butt, Carmel Delzoppo, Johnny Millar, Ben Gelbart, Felix Oberender, Subodh Ganu, Georgia Letton, Marino Festa, Gail Harper.

Author contributions

LJS designed the study and supervised all aspects of study conduct, analyses, and wrote the first draft of the manuscript with PG. WT, SB, JK and KG contributed to study design, reviewed analyses, and contributed to manuscript drafting. PG performed the analyses and contributed to manuscript drafting. DL, RV and JM contributed to study design and manuscript drafting. All authors have reviewed and approved the final manuscript.

Data sharing statement

The data are owned by the Australian and New Zealand Intensive Care Registry. The authors can be contacted for data inquiries.

Declarations

Conflicts of interest

None declared. The funding bodies had no role in study design, conduct, analyses and reporting.

Ethics approval and consent

The study was approved by the Human Research and Ethics Committee (Children`s Health Queensland, Brisbane, Australia; number HREC/16/QRCH/255), including waiver of consent.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Moynihan KM, Alexander PMA, Schlapbach LJ, Millar J, Jacobe S, Ravindranathan H, Croston EJ, Staffa SJ, Burns JP, Gelbart B, Australian, New Zealand Intensive Care Society Pediatric Study G, the ACfO, Resource E (2019) Epidemiology of childhood death in Australian and New Zealand intensive care units. Intensive Care Med 45:1262–1271
2. Schlapbach LJ, Straney L, Alexander J, MacLaren G, Festa M, Schibler A, Slater A, Group APS (2015) Mortality related to invasive infections, sepsis, and septic shock in critically ill children in Australia and New Zealand, 2002–13: a multicentre retrospective cohort study. Lancet Infect Dis 15:46–54
3. Schlapbach LJ, Gibbons KS, Horton SB, Johnson K, Long DA, Buckley DHF, Erickson S, Festa M, d'Udekem Y, Alphonso N, Winlaw DS, Delzoppo C, van Loon K, Jones M, Young PJ, Butt W, Schibler A, Nitric Study Group tA, New Zealand Intensive Care Society Clinical Trials G, the APSG (2022) Effect of nitric oxide via cardiopulmonary bypass on ventilator-free days in young children undergoing congenital heart disease surgery: the NITRIC Randomized Clinical Trial. JAMA 328:38–47
4. Carlton EF, Pinto N, Smith M, Fink EL, Watson RS, Hoehn KS, Marupudi N, Dervan LA, Lackey M, Ringwood M, Maddux AB, Network tP-PlotP, Health tEKSNIoC, Network HDCPCCR (2021) Overall health following pediatric critical illness: a scoping review of instruments and methodology. Pediatr Crit Care Med 22:1061–1071
5. Fink EL, Jarvis JM, Maddux AB, Pinto N, Galyean P, Olson LM, Zickmund S, Ringwood M, Sorenson S, Dean JM, Carcillo JA, Berg RA, Zuppa A, Pollack MM, Meert KL, Hall MW, Sapru A, McQuillen PS, Mourani PM, Watson RS, Pediatric Acute Lung I, Sepsis Investigators Long-term Outcomes Subgroup Investigators a, Eunice Kennedy Shriver National Institute of Child H, Human Development Collaborative Pediatric Critical Care Research N (2020) Development of a core outcome set for pediatric critical care outcomes research. Contemp Clin Trials 91:105968
6. Long D, Gibbons K, Le Brocque R, Schults JA, Kenardy J, Dow B (2022) Midazolam exposure in the paediatric intensive care unit predicts acute post-traumatic stress symptoms in children. Aust Crit Care 35:408–414
7. Raman S, Brown G, Long D, Gelbart B, Delzoppo C, Millar J, Erickson S, Festa M, Schlapbach LJ (2021) Priorities for paediatric critical care research: a modified Delphi study by the Australian and New Zealand Intensive Care Society Paediatric Study Group. Crit Care Resuscit Austral Acad Crit Care Med 23:194–201
8. Van De Sande D, Van Genderen ME, Huiskens J, Gommers D, Van Bommel J (2021) Moving from bytes to bedside: a systematic review on the use of artificial intelligence in the intensive care unit. Intensive Care Med 47:750–760
9. Sanchez-Pinto LN, Luo Y, Churpek MM (2018) Big data and data science in critical care. Chest 154:1239–1248
10. Schlapbach LJ, MacLaren G, Festa M, Alexander J, Erickson S, Beca J, Slater A, Schibler A, Pilcher D, Millar J, Straney L, Australian, New Zealand Intensive Care Society Centre for O, Resource E, Australian, New Zealand Intensive Care Society Paediatric Study G (2017) Prediction of pediatric sepsis mortality within 1 h of intensive care admission. Intensive Care Med 43:1085–1096
11. Collins GS, Reitsma JB, Altman DG, Moons K (2015) Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. BMC Med 13:1
12. Hanly M, Edwards B, Goldfeld S, Craven RG, Mooney J, Jorm L, Falster K (2019) School starting age and child development in a state-wide, population-level cohort of children in their first year of school in New South Wales, Australia. Early Childh Res Q 48:325–340
13. Lain SJ, Bentley JP, Wiley V, Roberts CL, Jack M, Wilcken B, Nassar N (2016) Association between borderline neonatal thyroid-stimulating hormone concentrations and educational and developmental outcomes: a population-based record-linkage study. Lancet Diabetes Endocrinol 4:756–765
14. Tomaszewski W, Ablaza C, Straney L, Taylor C, Millar J, Schlapbach LJ (2022) Educational outcomes of childhood survivors of critical illness—a population-based linkage study. Crit Care Med 50:901–912
15. Werner H, Latal B, Valsangiacomo Buechel E, Beck I, Landolt MA (2014) Health-related quality of life after open-heart surgery. J Pediatr 164(254–258):e251
16. Landolt MA, Vollrath M, Laimbacher J, Gnehm HE, Sennhauser FH (2005) Prospective study of posttraumatic stress disorder in parents of children with newly diagnosed type 1 diabetes. J Am Acad Child Adolesc Psychiatry 44:682–689
17. Schlapbach LJ, Aebischer M, Adams M, Natalucci G, Bonhoeffer J, Latzin P, Nelle M, Bucher HU, Latal B, Swiss Neonatal N, Follow-Up G (2011) Impact of sepsis on neurodevelopmental outcome in a Swiss National Cohort of extremely premature infants. Pediatrics 128:e348-357
18. Hastie T, Tibshirani R, Friedman JH, Friedman JH (2009) The elements of statistical learning: data mining, inference, and prediction. Springer, New York

19. Breiman L (2001) Random forests. Mach Learn 1:5–32
20. Meng L, Pammi M, Saronwala A, Magoulas P, Ghazi AR, Vetrini F, Zhang J, He W, Dharmadhikari AV, Qu C, Ward P, Braxton A, Narayanan S, Ge X, Tokita MJ, Santiago-Sim T, Dai H, Chiang T, Smith H, Azamian MS, Robak L, Bostwick BL, Schaaf CP, Potocki L, Scaglia F, Bacino CA, Hanchard NA, Wangler MF, Scott D, Brown C, Hu J, Belmont JW, Burrage LC, Graham BH, Sutton VR, Craigen WJ, Plon SE, Lupski JR, Beaudet AL, Gibbs RA, Muzny DM, Miller MJ, Wang X, Leduc MS, Xiao R, Liu P, Shaw C, Walkiewicz M, Bi W, Xia F, Lee B, Eng CM, Yang Y, Lalani SR (2017) Use of exome sequencing for infants in intensive care units: ascertainment of severe single-gene disorders and effect on medical management. JAMA Pediatr 171:e173438
21. Marsland S (2011) Machine learning: an algorithmic perspective. Chapman and Hall/CRC, Boca Raton
22. Lipkin PH, Macias MM (2020) Promoting optimal development: identifying infants and young children with developmental disorders through developmental surveillance and screening. Pediatrics 145(1):e20193449
23. Committee on Children with Disabilities (2001) Developmental surveillance and screening of infants and young children. Pediatrics 108:192–196
24. Lundberg SM, Lee SI (2017) A unified approach to interpreting model predictions. Adv Neural Inf Process Syst 30
25. Straney L, Clements A, Parslow RC, Pearson G, Shann F, Alexander J, Slater A, Group APS, the Paediatric Intensive Care Audit N (2013) Paediatric index of mortality 3: an updated model for predicting mortality in pediatric intensive care*. Pediatr Crit Care Med 14:673–681
26. de Jonge M, Wubben N, van Kaam CR, Frenzel T, Hoedemaekers CWE, Ambrogioni L, van der Hoeven JG, van den Boogaard M, Zegers M (2022) Optimizing an existing prediction model for quality of life one-year post-intensive care unit: an exploratory analysis. Acta Anaesthesiol Scand 66:1228–1236
27. Wubben N, van den Boogaard M, Ramjith J, Bisschops LLA, Frenzel T, van der Hoeven JG, Zegers M (2021) Development of a practically usable prediction model for quality of life of ICU survivors: a sub-analysis of the MONITOR-IC prospective cohort study. J Crit Care 65:76–83
28. Zimmerman JJ, Banks R, Berg RA, Zuppa A, Newth CJ, Wessel D, Pollack MM, Meert KL, Hall MW, Quasney M, Sapru A, Carcillo JA, McQuillen PS, Mourani PM, Wong H, Chima RS, Holubkov R, Coleman W, Sorenson S, Varni JW, McGalliard J, Haaland W, Whitlock K, Dean JM, Reeder RW, Evaluation LAPS, I, (2020) Critical illness factors associated with long-term mortality and health-related quality of life morbidity following community-acquired pediatric septic shock. Crit Care Med 48:319–328
29. Fink EL, Maddux AB, Pinto N, Sorenson S, Notterman D, Dean JM, Carcillo JA, Berg RA, Zuppa A, Pollack MM, Meert KL, Hall MW, Sapru A, McQuillen PS, Mourani PM, Wessel D, Amey D, Argent A, de Carvalho WB, Butt W, Choong K, Curley MAQ, Del PilarAriasLopez M, Demirkol D, Grosskreuz R, Houtrow AJ, Knoester H, Lee JH, Long D, Manning JC, Morrow B, Sankar J, Slomine BS, Smith M, Olson LM, Watson RS, Pediatric Outcomes SaPlot-PALI, Sepsis Investigators N, the Eunice Kennedy Shriver National Institute of Child H, Human Development Collaborative Pediatric Critical Care Research N (2020) A core outcome set for pediatric critical care. Crit Care Med 48:1819–1828
30. Smart D, Youssef GJ, Sanson A, Prior M, Toumbourou JW, Olsson CA (2017) Consequences of childhood reading difficulties and behaviour problems for educational achievement and employment in early adulthood. Br J Educ Psychol 87:288–308
31. Rosenbaum JE, Kariya T (1991) Do school achievements affect the early jobs of high school graduates in the United States and Japan? Sociol Educ 64:78–95
32. Adams EK, Hancock KJ, Taylor CL (2020) Student achievement against national minimum standards for reading and numeracy in Years 3, 5, 7 and 9: a regression discontinuity analysis. Aust J Soc Issues 55:275–301
33. Watson RS, Beers SR, Asaro LA, Burns C, Koh MJ, Perry MA, Angus DC, Wypij D, Curley MAQ, Ackerson J, Best M, Conder AA, Cvijanovich NZ, Dempsey R, Dowling MD, Duvall SW, Fabricatore A, Gopin CB, Harder LL, Herringshaw A, Herrington K, Hertzog JH, Hunter SJ, Ionascu N, Klein GE, Lauer RE, Lichtenstein JD, Manning MM, Martin R, Mayfield JW, Morse MM, Moss EM, Mucci GA, Ramirez A, Rice SA, Rodriquez-Rivera M, Salorio CF, Smith AD, Smith JT, Thomas NH, Thomas A, Timm JR, Vandetta L, Wallace M, Warschausky SA, Warzak WJ, White DA, Willen EJ, Woodcock-Burroughs LJ, Zayat MM (2022) Association of acute respiratory failure in early childhood with long-term neurocognitive outcomes. JAMA 327:836
34. Verlinden I, Güiza F, Dulfer K, Van Cleemput H, Wouters PJ, Guerra GG, Joosten KF, Verbruggen SC, Vanhorebeek I, Van den Berghe G (2022) Physical, emotional/behavioral, and neurocognitive developmental outcomes from 2 to 4 years after PICU admission: a secondary analysis of the early versus late parenteral nutrition randomized controlled trial cohort. Pediatr Crit Care Med 23:580–592
35. Mesotten D, Gielen M, Sterken C, Claessens K, Hermans G, Vlasselaers D, Lemiere J, Lagae L, Gewillig M, Eyskens B, Vanhorebeek I, Wouters PJ, Van den Berghe G (2012) Neurocognitive development of children 4 years after critical illness and treatment with tight glucose control: a randomized controlled trial. JAMA 308:1641–1650
36. Morris KP, Kapetanstraki M, Wilkins B, Slater AJ, Ward V, Parslow RC (2022) Lactate, base excess, and the pediatric index of mortality: exploratory study of an international, multicenter dataset. Pediatr Crit Care Med 23:e268–e276
37. Kyösti E, Peltoniemi O, Liisanantti JH, Ohtonen P, Ebeling H, Spalding M, Rautiainen P, Kataja J, Ala-Kokko TI (2020) School performance after pediatric intensive care-association of mental well-being, chronic illnesses, and family socioeconomic status. Pediatr Crit Care Med 21:e1099–e1105
38. Joseph RM, O'Shea TM, Allred EN, Heeren T, Hirtz D, Jara H, Leviton A, Kuban KC (2016) Neurocognitive and academic outcomes at age 10 years of extremely preterm newborns. Pediatrics 137
39. Laurin JC, Geoffroy MC, Boivin M, Japel C, Raynault MF, Tremblay RE, Cote SM (2015) Child care services, socioeconomic inequalities, and academic performance. Pediatrics 136:1112–1124
40. The L (2017) The health inequalities and ill-health of children in the UK. Lancet 389:477
41. Stringhini S, Carmeli C, Jokela M, Avendano M, Muennig P, Guida F, Ricceri F, d'Errico A, Barros H, Bochud M, Chadeau-Hyam M, Clavel-Chapelon F, Costa G, Delpierre C, Fraga S, Goldberg M, Giles GG, Krogh V, Kelly-Irving M, Layte R, Lasserre AM, Marmot MG, Preisig M, Shipley MJ, Vollenweider P, Zins M, Kawachi I, Steptoe A, Mackenbach JP, Vineis P, Kivimaki M, consortium L (2017) Socioeconomic status and the 25 x 25 risk factors as determinants of premature mortality: a multicohort study and meta-analysis of 1.7 million men and women. Lancet 389:1229–1237
42. Marmot M (2015) The health gap: the challenge of an unequal world. Lancet 386:2442–2444
43. Iqbal O'Meara AM, Miller Ferguson N, Zven SE, Karam OL, Meyer LC, Bigbee JW, Sato-Bigbee C (2020) Potential neurodevelopmental effects of pediatric intensive care sedation and analgesia: repetitive benzodiazepine and opioid exposure alters expression of glial and synaptic proteins in juvenile rats. Crit Care Explor 2:e0105
44. Laptook AR, Shankaran S, Tyson JE, Munoz B, Bell EF, Goldberg RN, Parikh NA, Ambalavanan N, Pedroza C, Pappas A, Das A, Chaudhary AS, Ehrenkranz RA, Hensman AM, Van Meurs KP, Chalak LF, Khan AM, Hamrick SEG, Sokol GM, Walsh MC, Poindexter BB, Faix RG, Watterberg KL, Frantz ID 3rd, Guillet R, Devaskar U, Truog WE, Chock VY, Wyckoff MH, McGowan EC, Carlton DP, Harmon HM, Brumbaugh JE, Cotten CM, Sanchez PJ, Hibbs AM, Higgins RD (2017) Effect of therapeutic hypothermia initiated after 6 hours of age on death or disability among newborns with hypoxic-ischemic encephalopathy: a randomized clinical trial. JAMA 318:1550–1560
45. Spittle A, Orton J, Anderson PJ, Boyd R, Doyle LW (2015) Early developmental intervention programmes provided post hospital discharge to prevent motor and cognitive impairment in preterm infants. Cochrane Database Syst Rev 2015:CD005495
46. Mitteregger E, Dirks T, Theiler M, Kretschmar O, Latal B (2022) A family-tailored early motor intervention (EMI-Heart) for infants with complex congenital heart disease: study protocol for a feasibility RCT. Pilot Feasibil Stud 8:263
47. Zendarski N, Sciberras E, Mensah F, Hiscock H (2017) Academic achievement and risk factors for adolescents with attention-deficit hyperactivity disorder in middle school and early high school. J Dev Behav Pediatr 38:358–368
48. Leder G, Lubienski S (2015) Large-scale test data: making the invisible visible. In: Bishop A, Tan H, Barkatsas TN (eds) Diversity in mathematics education: towards inclusive practices. Springer, Cham, pp 17–40
49. Lowrie T, Logan T, Ramful A (2016) Cross cultural comparison of grade 6 students' performance and strategy use on graphic and non-graphic tasks. Learn Individ Differ 52:97–108