Early detection of increased intracranial pressure episodes in traumatic brain injury: external validation in an adult and in a pediatric cohort

Fabian Güiza, PhD
Department of Intensive Care Medicine, University Hospitals Leuven, Leuven
Belgium

Bart Depreitere, MD, PhD
Department of Neurosurgery, University Hospitals Leuven, Leuven
Belgium

Ian Piper, PhD
Department of Clinical Physics, Southern General Hospital, Glasgow
UK

Giuseppe Citerio, MD
San Gerardo Hospital, Monza
Italy

Philippe G Jorens, MD, PhD
Department of Intensive Care Medicine, Antwerp University Hospital, Edegem
Belgium

Andrew Maas, MD, PhD
Department of Neurosurgery, Antwerp University Hospital, Edegem
Belgium

Martin U. Schuhmann, MD, PhD
Klinik für Neurochirurgie, Universitätsklinikum Tübingen, Tübingen
Germany

Tsz-Yan Milly Lo, MD
Department of Paediatric Intensive Care, Royal Hospital for Sick Children, Edinburgh
UK

Rob Donald, PhD
School of Mathematics and Statistics, University of Glasgow, Glasgow
UK

Patricia Jones, MD
Department of Paediatric Neurology, Royal Hospital for Sick Children, Edinburgh
UK

Gottlieb Maier, MD
Klinik für Neurochirurgie, Universitätsklinikum Tübingen, Tübingen,
Germany

Greet Van den Berghe, MD, PhD
Department of Intensive Care Medicine, University Hospitals Leuven, Leuven
Belgium

Geert Meyfroidt, MD, PhD
Department of Intensive Care Medicine, University Hospitals Leuven, Leuven
Belgium

Funding information:

Supported by the Foundation for Scientific Research Flanders (Fonds Wetenschappelijk Onderzoek - Vlaanderen [FWO]) (research project G. 0904.11). Brain-IT was funded by the European Framework Programme (FP5-QLRI-2000-00454, QLGT-2002-00160, and FP7-IST-2007–217049). The NEMO project in the University Hospital Edegem (Antwerp), Belgium, was funded by the Flemish Government Agency for Innovation by Science and Technology (IWT)—Applied Biomedical Research (TBM) program. Dr. Depreitere's institution received funding from the Johnson&Johnson educational grant. Dr. Piper received support for article research (EU Grants funded data collection of some of the data [Brain-IT data] upon which the project analyses were performed. FP5 QLGT-2002-00160 and FP7 IST-2007–217049). Dr. Citerio's institution received funding from Educational activities for Codman. Dr. Maas' institution received funding from the Framework 7 Program European Union and from IWT (Flemish agency for Innovation on Science and Technology). Dr. Jones received support for article research from Chief Scientist Office, Scotland, UK Brain-IT. Dr. Van den Berghe received long-term structural research financing via the Methusalem program funded by the Flemish Government (METH/08/07). Dr. Meyfroidt received funding from FWO as senior clinical investigator (1846113N). The remaining authors have disclosed that they do not have any potential conflicts of interest.

Corresponding author contact information:

Fabian Güiza
Department of Intensive Care Medicine,
University Hospitals Leuven, Herestraat 49,
3000 Leuven, Belgium
e-mail: fabian.guiza@kuleuven.be
Phone +32 16 344021
Fax +32 16 344015

Key Words:

automated; data mining; decision support techniques; forecasting; intensive care; intracranial hypertension; models; pattern recognition; statistical; traumatic brain injury

Objective: A model for early detection of episodes of increased intracranial pressure in traumatic brain injury patients has been previously developed and validated based on retrospective adult patient data from the multicenter Brain-IT database. The purpose of the present study is to validate this early detection model in different cohorts of recently treated adult and pediatric traumatic brain injury patients.

Design: Prognostic modelling. Noninterventional, observational, retrospective study.

Setting and Patients: The adult validation cohort is comprised of recent traumatic brain injury patients from San Gerardo Hospital in Monza (n=50), Leuven University Hospital (n=26), Antwerp University Hospital (n=19), Tübingen University Hospital (n=18) and Southern General Hospital in Glasgow (n=8). The pediatric validation cohort is comprised of patients from neurosurgical and intensive care centres in Edinburgh and Newcastle (n=79).

Interventions: None

Measurements and Main Results: The model's performance was evaluated with respect to discrimination, calibration, overall performance, and clinical usefulness. In the recent adult validation cohort, the model retained excellent performance as in the original study. In the pediatric validation cohort, the model retained good discrimination and a positive net benefit, albeit with a performance drop in the remaining criteria.

Conclusions: The obtained external validation results confirm the robustness of the model to predict future increased intracranial pressure events 30 minutes in advance, in adult and pediatric traumatic brain injury patients. These results are a large step towards an early warning system for increased intracranial pressure that can be generally applied. Furthermore, the sparseness of this model that uses only two routinely monitored signals as inputs (intracranial pressure and mean arterial blood pressure) is an additional asset.

Introduction

Elevations in intracranial pressure (ICP) are considered to be life- and brain-threatening secondary injuries after severe traumatic brain injuries (TBI) [1]. Management of intracranial hypertension is mainly reactive, with increased ICP episodes treated aggressively. Early warnings could improve proactive management by increasing the time window for diagnostic and therapeutic measures. To that end, a model to detect future increased ICP episodes 30 minutes in advance, by analyzing the characteristics of the four previous hours of the ICP and mean arterial blood pressure (MAP) signals, was developed and validated in a large international multicenter dataset of 178 patients of the EU-funded Brain Monitoring with Information Technology (Brain-IT) project [2]. This model was able to predict increased ICP episodes with good discrimination, calibration and overall model performance [3]. The predictive power was found predominantly in the ICP signal itself, with the most recent measurements being more relevant. Also predictive was the correlation between ICP and MAP, which hints at the prognostic importance of cerebrovascular autoregulation.

In the present study, this model was externally validated in two new independent datasets: an adult and a pediatric cohort of TBI patients. This work has been presented in part, at the 2014 International Symposium on Intensive Care and Emergency Medicine (Brussels).

Materials and Methods

External Validation Databases

The adult cohort comprises 121 adult patients from five European centers: 50 admitted to the ICU of the San Gerardo Hospital in Monza, Italy, between March 2010 and April 2013; 26 admitted to the ICU's of the University Hospitals Leuven, Belgium, between September 2010 and September 2013; 19 admitted to the ICU of the University Hospital Edegem (Antwerp), Belgium as part of the 'Individualized targeted monitoring in neurocritical care' (NEMO) project between May 2010 and June 2013 [4]; 18 admitted to the ICU of the University Hospital Tübingen, Germany, between February and December 2009; and eight admitted to the ICU of the Southern General Hospital, Glasgow between September 2009 and July 2011. All data pertaining to patient identity were removed, and ethical committee approval was obtained from all centers for retrospective data analysis. In Antwerp, informed consent was granted by the closest relatives prior to enrollment. Monitoring data for the NEMO patients was recorded every second; the median value per minute was used in this study. There were 231 instances of 4-hour time

series of ICP and MAP leading to increased ICP episodes within the 30-minute time horizon, with at least one episode occurring in 41 patients; or equivalently an approximate average instance-to-patient ratio of 6:1. Additional 820 instances of 4-hour time series that did not precede increased ICP episodes within the 30-minute horizon were randomly selected from all available instances from the 121 patients, thus preserving the average instance-to-patient ratio.

The pediatric cohort comprises 79 children recruited during 62 nonconsecutive months up to July 2003 from two regional pediatric neurosurgical and intensive care centers in Edinburgh and Newcastle, United Kingdom [5]. The study had local ethics committee and management approval in both centers, and informed consent was obtained before enrollment in the study. There were 811 instances preceding increased ICP episodes, with at least one episode occurring in 49 patients. To preserve the average instance-to-patient ratio in this case, additional 1408 instances not preceding hypertension episodes were randomly sampled from all 79 patients.

In all cases, minute-by-minute signals were reviewed independently by two clinicians in Leuven (G.M., B.D.), and obvious artifacts through visual inspection were excluded from further analysis. **Table 1** summarizes the cohorts' demographic information. As in the study of Güiza et al [3], an increased ICP episode was defined as above 30 mm Hg and lasting at least 10 consecutive minutes.

Evaluation Criteria

Discrimination between events and nonevents of increased ICP was evaluated visually with box plots and numerically with the area under the receiver operating characteristic (AUROC) curve, and with the discrimination slope (DS), which is the absolute difference between average predictions for events and nonevents. Calibration (the agreement between observed outcomes and predictions) was evaluated visually with calibration plots and numerically with the Hosmer–Lemeshow (HL) statistic, calibration–in–the–large, and the calibration slope. Overall model performance was evaluated with the Brier score (BS) and the Brier score scaled (BSS). Sensitivity, specificity, accuracy, true positives (TP) and false positives (FP) were evaluated at different cutoff probabilities. The clinical usefulness or net benefit (NB) from making treatment decisions based on the predictions was evaluated visually with decision curves. A detailed description of these criteria can be found in the study by Steyerberg et al [6]. All analyses were done in MATLAB 2014b (The MathWorks, Natick, MA).

Results

Results are reported in **Table 2** and **Figure 1**. Results for the original study are reproduced in the first column, the second and last columns report performance in the adult and pediatric cohorts, respectively.

In the adult cohort, good discrimination is clear from the box plots' separation between event and nonevent predictions, a 0.90 AUROC and a 0.39 DS. The model remains well calibrated, with a HL *p* value above 0.05, calibration-in-the-large close to 0, calibration-slope close to 1, and a calibration curve close to the diagonal. Overall model performance remains with 0.10 BS and 40% BSS. At the same probability cutoff of the original study, performance deviates to higher accuracy (86%), lower sensitivity (70%), and higher specificity (90%). At this cutoff, there is a 59% reduction in FP when compared with the alert-all policy and a 11% increase in TP when compared with the alert-none policy. The decision curve demonstrates positive NB above default policies for all risk thresholds.

In the pediatric cohort, good discrimination is evidenced by the clear box plots' separation, a 0.79 AUROC and a 0.28 DS. However, predictions are systematically high as observed in the box plots, the deviation from the diagonal in the calibration curve, a -0.14 calibration-in-the-large, a 0.80 calibration slope and a HL *p* value less than 0.05. This has a detrimental effect on overall performance resulting in 0.2 BS and 15% BSS. At the same probability cutoff of the original study, performance deviates to higher sensitivity (92%), lower specificity (48%), and lower accuracy (64%). There is a 25.2% reduction in FP when compared with alert-all and a 14.2% increase in TP when compared with alert-none. The decision curve shows NB above default policies only for risk thresholds lower than 0.6.

Discussion

Performance as determined by AUROC and DS in both adult and pediatric validation cohorts confirm the model's capability to discriminate the onset of an increased ICP episode 30 minutes in advance, irrespective of age. Confirmatory studies in general are an important and mandatory step of the scientific process, in their own right adding to the general knowledge in the field [7]. This study is particularly relevant for neurocritical care, as it first shows the model, developed using data prior to 2005, can still discriminate well in adult patients treated after the 2007 change in the Brain Trauma Foundation guidelines [8]. Second, despite a decrease in discrimination for the pediatric cohort, the model remains clinically useful; which is a relevant finding as only 7.3% of the patients used for model development in

the original study [3] were children; and furthermore to the best of our knowledge, this is the first time early detection of increased ICP has been studied in a pediatric setting. This discriminatory capacity across cohorts is likely due to the model's sparseness, as it only uses information derived from two routinely monitored signals (ICP and MAP).

Beyond discrimination, a model's usefulness increases when it is well calibrated [9]. In this regard, the model only complies for the adult cohort. In practice, however, clinical usefulness or NB is determined by performance at a chosen risk threshold, i.e., the probability cutoff chosen to decide for treatment [10]. Treatment here refers to preparatory actions for an upcoming hypertension event, such as stricter observation or possible transport to imaging. The costs and harm associated with treatment are center and patient specific. However, because these actions will arguably pose little harm to the patient, risk thresholds below 0.5 are recommended [10]. The decision curves show that within this threshold range, the model provides higher NBs than default policies in all cohorts, including the pediatric case. Noticeably so, compared with current practice in which no warning exists (alert-none policy).

When computing each prediction, the model assigns more weight to previously seen instances that are most similar to the one under evaluation. Such similarity-based approaches have also proven successful in other TBI prediction settings [11]. The fact that there were few children in the original dataset, used for model development, partially explains the decrease in model performance in the pediatric validation cohort.

Limitations of this study are its retrospective nature and the relatively few patients per cohort. Regrettably, storage of minute-by-minute monitoring data in TBI is not common, and consequently few databases of this sort exist. Nevertheless, the number of elevated ICP episodes studied per cohort was sufficient for external validation [12]. Although it does not affect its intended use as an early warning system, a possible shortcoming of the model is that it does not provide an explanation for the underlying physiological mechanisms driving the impending elevated ICP episode.

Conclusion

The external validation performed in two independent datasets is a large step towards an early warning system for increased ICP episodes in TBI that can be used in ICUs worldwide. Because it is based only on two routinely monitored signals, our model appears to be generally applicable for use in

TBI patients. Model performance remains unchanged in a cohort of recently treated adult patients and has positive NB when used for clinical decision making irrespective of age. Future intervention studies are required to assess the impact of this model on patient outcome when used in clinical practice.

Acknowledgments

We wish to acknowledge the non-co-author members of the Brain-IT steering group: Iain Chambers, Per Enblad, Barbara Gregson, Tim Howells, Karl Kiening, Julia Mattern, Pelle Nillson, Arminas Ragauskas, and Juan Sahuquillo. We would also like to acknowledge the contributors to the original study resulting in the pediatric database used here: R.J. Forsyth, B. Fulton, P.J.D. Andrews, A.D. Mendelow, I.R. Chambers, and R.A. Minns.

References:

[1] Jones PA, Andrews PJD, Midgley S, et al: Measuring the burden of secondary insults in head injured patients during intensive care. J Neurosurg Anaesth 1994; 6:4–14

[2] Piper I, Citerio G, Chambers I, et al. The BrainIT group: Concept and core dataset definition. Acta Neurochirurgica (Wien) 2003; 145: 615–28; discussion 628

[3] Güiza F, Depreitere B, Piper I, et al: Novel methods to predict increased intracranial pressure during intensive care and long-term neurological outcome after traumatic brain injury: Development and validation in a multicenter dataset. Crit Care Med 2013; 41:554-564

[4] Feyen BF, Sener S, Jorens PG, et al: Neuromonitoring in traumatic brain injury. Minerva Anestesiol 2012; 78: 949-58

[5] Chambers IR, Jones PA, Lo TY, et al: Critical thresholds of intracranial pressure and cerebral perfusion pressure related to age in paediatric head injury. J Neurol Neurosurg Psychiatry 2006; 77:234-240.

[6] Steyerberg EW, Vickers AJ, Cook NR, et al: Assessing the performance of prediction models: A framework for traditional and novel measures. Epidemiology 2010; 21:128-138.

[7] Ioannidis JP: Why most published research findings are false. PLoS Med 2005; 2:e124

[8] Brain Trauma Foundation; American Association of Neurological Surgeons; Congress of Neurological Surgeons; Joint Section on Neurotrauma and Critical Care, AANS/CNS, Bratton SL,

Chestnut RM, Ghajar J, et al: Guidelines for the management of severe traumatic brain injury. VIII. Intracranial pressure thresholds. J Neurotrauma 2007; 24:S55–S58

[9] Steyerberg EW: Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. New York, Springer, 2009; XXVIII, 500 p.

[10] Vickers AJ, Elkin EB: Decision curve analysis: A novel method for evaluating prediction models. Med Decis Making 2006; 26:565–74.

[11] Nathanson BH, Higgins TL, Giglio RJ, et al: An exploratory study using data envelopment analysis to assess neurotrauma patients in the intensive care unit. Health Care Manag Sci. 2003; 6:43-55

[12] Peek N, Arts DG, Bosman RJ, et al: External validation of prognostic models for critically ill patients required substantial sample sizes. J Clin Epidemiol 2007; 60:491-501.

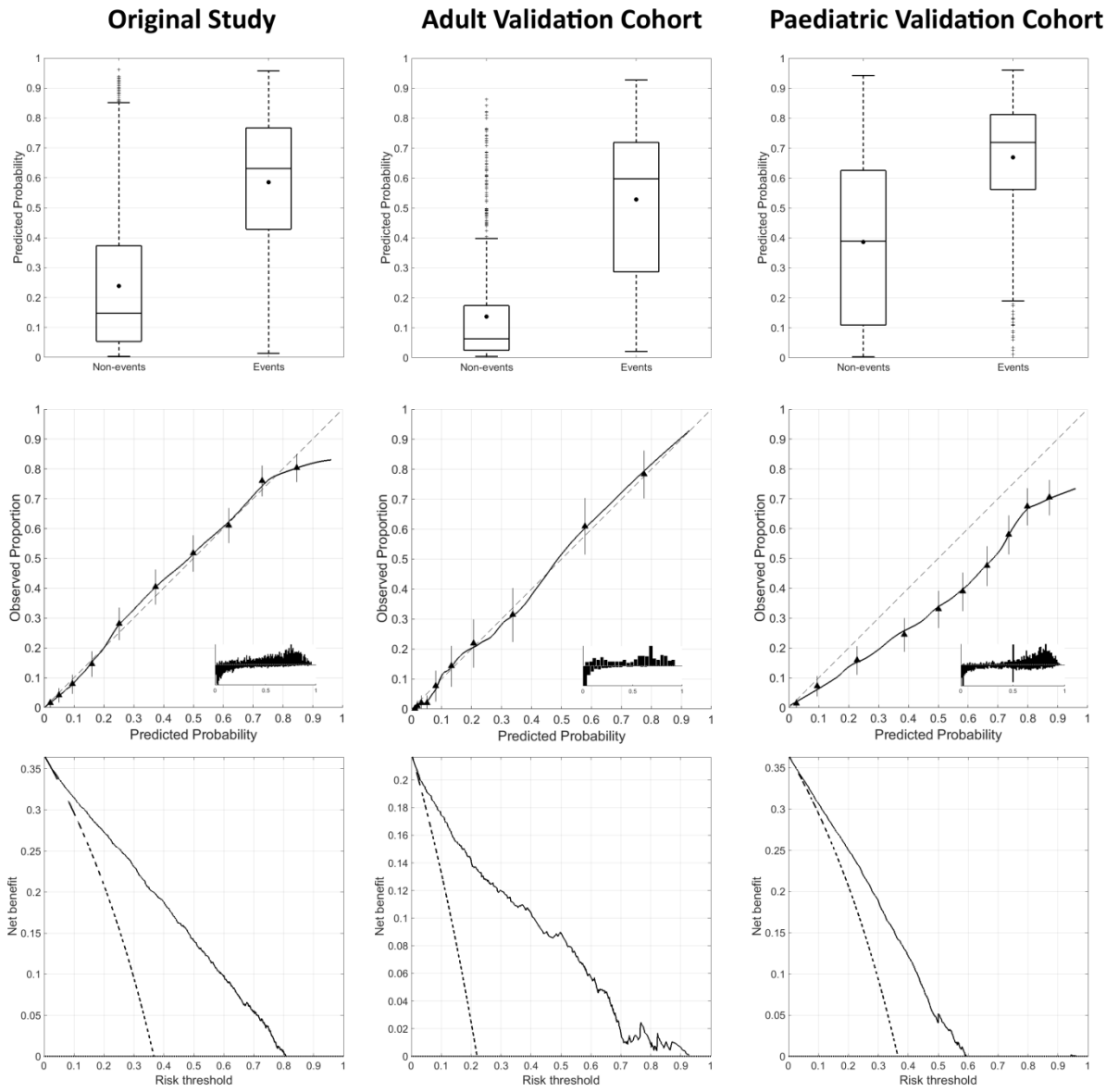

Figure 1. The first column corresponds to the Brain-IT development cohort of the original study; the second column to the recent adult validation cohort and the third column to the pediatric validation cohort. The first row depicts box plots of predicted probabilities for nonevents versus events of increased intracranial pressure (the *dot* shows the mean of each distribution). The second row depicts calibration plots (observed proportion of events by deciles of predicted probabilities). The *dashed diagonal line* depicts perfect calibration and the *solid line* the curve obtained with Locally Weighted Smoothing Scatterplot smoothing. Distributions of predictions for events and nonevents are shown in histograms. The third row depicts decision curves, the *solid line* is the net benefit of treating instances with predicted probabilities greater than the risk threshold (x-axis), the *dashed line* is the net benefit of always treating (alert-all policy), while never treating (alert-none policy) has a constant net benefit of 0. Alert-all and alert-none intersect at the dataset prevalence.

Cohorts	**Original Study**	**External Validation**	
	Brain-IT Development	Adult Cohort	Pediatric Cohort
Number of patients (n)	178	121	79
Length of stay (d), median (25^{th} - 75^{th} percentile)	14 (7-23)	17 (8.8-26)	3 (2-6)
Age (yr), median (25^{th} - 75^{th} percentile)	33 (19-49)	50 (28.5-65)	10.4(7.5-14.2)
Sex (% men)	81	78	74
Pupil reactivity (%)			
None	12.4	6.6	7.6
One	7.3	9.1	13.9
Two	71.3	59.5	69.6
Unknown, untestable or missing	9	24.8	8.9
Glasgow Coma Scale total, median (25^{th} - 75^{th} percentile)	7 (4-10)	7 (3-12)	6(5-8)
Unknown, untestable or missing (%)	5.6	25.6	0.0
Increased ICP episodes per patient, median (25^{th} - 75^{th} percentile)	1 (0-6)	0(0-1)	1(0-8)
Increased ICP episodes per patient with at least one episode, median (25^{th} – 75^{th} percentile)	5 (1.5-12)	3 (1-7)	7.5 (1.5-21)

Table 1. Patient Demographic Information For All Cohorts. ICP = intracranial pressure. An episode of increased ICP was defined as an ICP above 30 mmHg lasting at least 10 consecutive minutes as described in the study by Güiza et al [3].

Cohorts	Original Study	External Validation	
	Brain-IT Development	Adult Cohort	Paediatric Cohort
Number of patients (n)	n=178	n=121	n=79
Patients (%) with at least one high ICP episode	61	34	62
Number of high ICP episodes (e)	982	231	811
Number of instances (i)	2677	1051	2219
Episode Prevalence (P) (%)	37	22	37
AUROC (C statistic)	0.85 [0.84–0.86]	0.90 [0.87–0.91]	0.79 [0.77–0.81]
Discrimination slope	0.35 [0.33–0.37]	0.39 [0.35–0.42]	0.29 [0.26–0.30]
Hosmer Lemeshow *p* value	0.18 [0.09–0.97]	0.70 [0.77–0.92]	<0.01 [<0.01–<0.01]
Calibration-in-the-large	0.00 [-0.01–0.01]	-0.003 [-0.02–0.01]	-0.14 [-0.14– -0.10]
Calibration-slope	0.99 [0.94–1.04]	1.04 [0.95–1.12]	0.81 [0.76–0.86]
Brier Score	0.15 [0.14–0.16]	0.10 [0.09–0.11]	0.20 [0.19–0.20]
BrierScore Scaled (%)	34 [31–37]	40 [34–46]	15 [10–19]
Accuracy (%)	77 [75–78]	86 [84–88]	64 [62–66]
Sensitivity (%)	79 [77–82]	70 [64–76]	91 [90–93]
Specificity (%)	76 [73–77]	90 [88–92]	48 [45–51]
False-positives reduction (w.r.t. alert-all) (%)	35 [32–37]	59 [55–64]	25 [23–27]
True-positives increase (w.r.t. alert-none) (%)	20 [18–22]	11 [8.4–14]	14 [11-16]

Table 2. Performance of the Model For Early Detection of Increased Intracranial Pressure Episodes. AUROC = area under the receiver operating characteristic, ICP = intracranial pressure. The last five rows were computed using the probability cutoff that maximized sensitivity and specificity in the development cohort. 95% confidence intervals obtained from 2000 bootstrap replicas are shown between square brackets.

Brier_max = prevalence × (1 – prevalence)^2+ (1 – prevalence) × prevalence^2.

Brier_max (original study) = 0.23; Brier_max (adult validation cohort) = 0.17; Brier_max (paediatric validation cohort) = 0.23;
Brier score scaled = 1 – (Brier score/Brier_max).