Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Development of ePILOT, an Electronic Medical Record–Based Measure of Intracranial Pressure–Directed Therapy Intensity in Pediatric Traumatic Brain Injury

Mohammed Altaf Shaik

shaikm3@upmc.edu

Children's Hospital of Pittsburgh of UPMC https://orcid.org/0000-0002-3667-247X

Jaskaran Rakkar

Phoenix Children's Hospital

Jonathan H. Pelletier

Akron Children's Hospital

Alicia K. Au

University of Pittsburgh School of Medicine

Michael J. Bell

The University of Texas Southwestern Medical Center

Dennis W. Simon

University of Pittsburgh School of Medicine

Patrick M. Kochanek

University of Pittsburgh School of Medicine

Gilles Clermont

University of Pittsburgh School of Medicine

Christopher Horvat

University of Pittsburgh School of Medicine

Robert S. B. Clark

University of Pittsburgh School of Medicine

Research Article

Keywords:

Posted Date: November 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11555789/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Development of ePILOT, an Electronic Medical Record–Based Measure of Intracranial Pressure–Directed Therapy Intensity in Pediatric Traumatic Brain Injury

Mohammed A. Shaik, MD, PhD,[1,2] Jaskaran Rakkar, MD,[3*] Jonathan H. Pelletier, MD, MS,[4] Alicia K. Au, MD, MS,[1,2,5] Michael J. Bell, MD,[6] Dennis W. Simon, MD,[1,2,5] Patrick M. Kochanek, MD,[1,2,5] Gilles Clermont, MD,[1,7] Christopher Horvat, MD, MHA,[1,2,5,7,8] Robert S. B. Clark, MD[1,2,5]*

Affiliations:

1. Department of Critical Care Medicine, University of Pittsburgh School of Medicine; Pittsburgh, PA, USA.
2. Safar Center for Resuscitation Research, University of Pittsburgh School of Medicine; Pittsburgh, PA, USA.
3. Division of Pediatric Critical Care, Phoenix Children's, Phoenix, AZ, USA.
4. Division of Critical Care, Department of Pediatrics, Akron Children's Hospital, Akron, OH, USA.
5. Department of Pediatrics, University of Pittsburgh School of Medicine; Pittsburgh, PA, USA.
6. Division of Pediatric Critical Care, UT Southwestern Medical Center, Dallas, TX, USA.
7. The Clinical Research, Investigation, and Systems Modeling of Acute Illness (CRISMA) Center, Pittsburgh, PA, USA.
8. Department of Pediatrics, Division of Health Informatics, UPMC Children's Hospital of Pittsburgh; Pittsburgh, PA, USA.

***MAS and JR contributed equally to this work.**

Address Correspondence To: Mohammed A. Shaik, Department of Critical Care Medicine, University of Pittsburgh School of Medicine, UPMC Children's Hospital of Pittsburgh, 4401 Penn Avenue, Pittsburgh, PA 15224, USA, shaikm3@upmc.edu, 917-741-8111

Support: 5T32HD040686-20 (JR, MAS, JHP)

Abstract: 299 words | **Manuscript:** 4370 words | **Figures:** 6 | **Tables:** 3 | **Supplemental:** 1

figure, 8 tables

DETAILS PAGE

1. Compliance with instructions. This manuscript complies with the Neurocritical Care Instructions for Authors.

2. Author contributions. MAS and JR contributed equally and are co-primary authors. MAS curated the data, performed the chart review that defined the analytic cohort, conducted the re-analysis and statistical analyses, created the exhibits, and drafted and revised the manuscript. JR conceived and designed ePILOT, acquired the data, performed the original analysis, and drafted the manuscript. CH and RSBC supervised the work. JHP, AKA, MJB, DWS, PMK, GC, CH and RSBC interpreted the data and critically revised the manuscript for important intellectual content.

3. Authorship. All authors meet the ICMJE criteria for authorship, and all authors read and approved the final manuscript.

4. Exclusive submission. This manuscript has not been published elsewhere and is not under consideration by another journal.

5. Ethics. The study was approved by the University of Pittsburgh Institutional Review Board, which waived the requirement for informed consent for this retrospective study. The study was conducted in accordance with the Declaration of Helsinki.

6. Conflicts of interest. The authors declare no competing interests.

7. Reporting checklists. This study is reported in accordance with STROBE (observational cohort) and TRIPOD+AI (derivation and evaluation of the score). The completed checklists are included at the end of the manuscript.

8. Funding. Supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) training grant 5T32HD040686-20 (JR, MAS, JHP).

9. Use of AI-assisted technology. Claude (Anthropic), a large language model, was used to write R/Python code for analysis and generation of tables and figures, to edit manuscript text, including reconciling co-author edits. and to check and format references against PubMed. The authors reviewed and verified all AI-assisted code, output and text and take full responsibility for the content.

ABSTRACT

Background/Objective: Existing therapy-intensity scales for traumatic brain injury (TBI) require manual abstraction and are scored at coarse intervals, limiting temporal alignment with intracranial pressure (ICP). We aimed to develop an hourly, electronic medical record (EMR)-derived Pediatric Intensity Level of Therapy (PILOT) scale (ePILOT) and characterize its relationships with ICP and in-hospital mortality.

Methods: We retrospectively analyzed hourly EMR data from the first 7 days of hospitalization in children with TBI who underwent ICP monitoring at a single center. Twelve components adapted from PILOT (cerebrospinal fluid diversion, hyperosmolar therapy, sedation/analgesia, barbiturates, neuromuscular blockade, temperature, $PaCO_2$, and serum sodium) were derived from routinely stored data. A Mixed Graphical Model descriptively characterized conditional dependencies among components and ICP. Integer component weights were informed by associations with in-hospital mortality and adjusted to preserve guideline tier ordering, yielding an hourly score of 0–24. Patient-level relationships with ICP and mortality were assessed.

Results: Among 135 ICP-monitored TBI patients (median age 4.8 [1.9–12.7] years; 65% male; admission Glasgow Coma Scale (GCS) score 4 [3–7]; mortality 16%), ePILOT increased progressively with ICP (patient-level Spearman $\rho = 0.32$, $p < 0.001$). Non-survivors had higher mean ICP (26 vs. 12 mmHg) and median hourly ePILOT (7.0 [5.0–10.4] vs. 3.5 [3.0–6.5]) than survivors (both $p < 0.05$). Peak hourly ePILOT discriminated mortality with an area under the receiver operating characteristic curve (AUROC) of 0.74 (95% CI 0.61–0.85), compared with 0.68–0.72 for ICP summaries and 0.61 for admission GCS. A median hourly ePILOT score of 5 was the Youden-optimal mortality cut-point (95% CI 4–8; sensitivity 0.77; specificity 0.61).

Conclusions: ePILOT converts routinely stored EMR data into an automated, hourly measure of ICP-directed therapy intensity. Therapy intensity tracked ICP, lagged behind it, and discriminated mortality comparably to ICP, consistent with a reactive treatment response. External validation against manually scored PILOT should follow.

INTRODUCTION

Traumatic brain injury (TBI) is a leading cause of mortality and morbidity in both children and adults, accounting for approximately 30% of all injury-related deaths in the United States (U.S.)[1]. In severe TBI, management centers on hemodynamic support and on invasive intracranial pressure (ICP) monitoring to guide therapy. While ICP-guided therapy has not been formally validated in randomized trials, its use has been associated with improved outcomes[2,3]. Fittingly, ICP-directed therapies remain a hallmark of TBI management under established guidelines[4] and the treatment of severe TBI follows a tiered, patient-centric approach formalized in consensus algorithms for both adults[5] and children[4].

Intracranial hypertension has been defined as a condition of increased ICP, with ICP > 20 mmHg often cited as the threshold in neurotrauma. Emerging pediatric data suggest that a lower threshold of 14–15 mmHg better discriminates outcome and may be more appropriate for children with TBI[6,7]. Given the association between intracranial hypertension and poor outcomes, in tiered ICU management, clinicians progressively intensify and combine ICP-directed therapies until ICP targets are achieved, as recommended by the 2019 pediatric consensus and guidelines-based algorithm[4]. However, the overall magnitude of that therapeutic effort remains poorly quantified. Differentiating the magnitude of therapeutic support required for ICP- and/or cerebral perfusion pressure (CPP)-directed management may therefore reveal a patient's cerebral pathophysiological status beyond the therapeutic targets themselves. With the dynamic needs imposed by evolving intracranial pathology after TBI, a standardized, granular assessment of therapeutic intensity may better capture the treatment actually delivered (as compared to ICP itself), while paving the way for automated clinical decision support for ICP/CPP-directed therapies[8].

For pediatric TBI, existing metrics include the Therapy Intensity Level (TIL) scale from 1987 and the Pediatric Intensity Level of Therapy (PILOT) scale developed in 2006, both parsimonious measures of the intensity of ICP-directed interventions[9,10]. The original TIL carried a barbiturate-related ceiling effect and omitted therapies such as hypertonic saline and temperature control, but was later redefined by inter-agency consensus into the current 38-point scale, which has since been clinimetrically validated and evaluated at scale within the multicenter CENTER-TBI cohort[11–13]. Both, however, rely on manual abstraction and are scored no more often than every 4 hours (TIL) or once daily (PILOT). Even where cerebral physiology is monitored minute-by-minute, therapeutic intensity is typically summarized only daily[14], a significant mismatch with how rapidly intracranial pathophysiology can evolve. To our knowledge, no therapy-intensity measure has been derived directly from routinely stored electronic medical record (EMR) data that is reproducible, scored hourly, and requires minimal additional data collection.

The data required by such a tool (ICP, CPP, arterial blood gases, serum chemistry, temperature, and medication administrations) are recorded in the EMR at least hourly in most pediatric intensive care units (PICUs), offering an opportunity to build it on existing infrastructure. Accordingly, we aimed to develop an EMR-based PILOT (ePILOT) scale to quantify therapeutic support used for ICP-directed care and to characterize the relationships between ICP, ICP-directed medical therapeutic support, and in-hospital mortality. We hypothesized that patients with a greater burden of intracranial hypertension receive more ICP-directed medical therapies and are at increased risk of mortality.

METHODS

Study Population and Data Collection

This study adheres to applicable EQUATOR guidelines, specifically STROBE[15] for the observational cohort and TRIPOD (TRIPOD+AI) for the reporting of the score's derivation and evaluation[16,17]. The study was conducted using PICU data extracted from the EMR at the UPMC Children's Hospital of Pittsburgh, a level I trauma center that houses a 36-bed PICU with a catchment area of approximately 5 million people. Eligibility criteria included patients who were admitted with TBI, were less than 18 years of age and had invasive ICP monitors placed. Of 461 patients admitted with invasive ICP monitoring during the study period (January 2009 – April 2019), 135 met eligibility criteria. Neonates were also included in this analysis due to an institutional policy mandating admission of all TBI patients to the PICU. The study was approved by the University of Pittsburgh Institutional Review Board, which waived the requirement for informed consent given the retrospective use of de-identified data. The study was conducted in accordance with the Declaration of Helsinki.

Data were collected from an enterprise data warehouse using SAP Business Objects (SAP, Walldorf, Germany) at an hourly frequency from admission for up to 7 days. The demographic and clinical data collected included age, sex, race, admission and discharge times, admission GCS score, hourly ICP values, discharge disposition, and in-hospital mortality. Additionally, the admission Pediatric Logistic Organ Dysfunction-2 (PELOD-2) score, a descriptive scoring system used to evaluate the severity of organ dysfunction, was computed for each patient. All data collected were de-identified prior to storage and analysis. Data elements were mapped to standard vocabularies (LOINC for measurements; RxNorm for medications; SNOMED CT for procedures) within the Observational Medical Outcomes Partnership (OMOP)

common data model[18], yielding a reproducible value set for external-site extraction (**Supplemental Table 1**).

Patient Management and Data Selection

Patient management adhered to the published TBI management guidelines[4,19–21]. The standard management protocol for severe pediatric TBI at our institution involves implementing a range of interventions, including the maintenance of adequate analgesia and sedation, neuromuscular blockade (NMB), and ensuring that PaO_2 is maintained between 100–150 mmHg (in the absence of brain tissue PO_2 monitoring) and $PaCO_2$ between 32–35 mmHg.

We developed ePILOT by adapting variables from the PILOT scale[10], which is composed of guideline-based treatments and interventions used to treat intracranial hypertension, defined as sustained ICP > 20 mmHg beyond 5 minutes[4,19–21]. Specific data elements included cerebrospinal fluid (CSF) diversion, osmotherapies[22], sedation/analgesia, NMB, barbiturates, $PaCO_2$, hypothermia, and serum sodium concentration. CSF diversion was scored in two parts: an external ventricular drain (EVD) in place earned 2 points, and active drainage (> 1 mL/h) earned an additional point, such that draining CSF to lower ICP scored as an escalation beyond simply having the EVD in place. Osmotherapies comprised bolus administration of 3% hypertonic saline (HTS) or mannitol, or a continuous infusion of either therapy. To assess the role of analgesia/sedation, we considered any additional bolus therapy of medications including dexmedetomidine, diazepam, etomidate, fentanyl, hydromorphone, ketamine, lorazepam, methadone, midazolam, morphine, pentobarbital and propofol. Pentobarbital infusion and NMB were evaluated separately. Disease-targeted mechanical ventilation settings were evaluated by measuring $PaCO_2$ on serial arterial blood gas samples. Continuous patient temperature was recorded by either esophageal or rectal temperature probes, and serum sodium levels were

included as an additional measure of the use of HTS, with the noted caveat of possible confounding for patients in diabetes insipidus (DI).

Thresholds for continuous variables, including $PaCO_2$, serum sodium and temperature, were selected a priori from published literature to distinguish routine therapeutic targets from physiologic extremes associated with treatment escalation or increased risk of complications[4,10]. The frequency of laboratory data varied based on the acuity of presentation and time since admission. Data imputation was performed to allow for standardization and inclusion of points when a value was not available. Imputation was conducted by carrying the last observation forward (LOCF) for at most a 4-hour interval, and only for $PaCO_2$, serum sodium, and temperature. When no new value had been recorded within 4 hours, the component scored 0. Medication, EVD and drainage components were scored in the hour in which they were documented and were not carried forward. We explicitly did not include surgical interventions such as decompressive craniectomy (DC) in ePILOT as these interventions were performed at the discretion of the pediatric neurosurgery team. Our efforts focused solely on intensivist-driven medical therapies at the bedside. DC at any time during the admission was identified by chart review, but exact timing was difficult to discern and therefore not recorded.

Statistical Methods

We used Mixed Graphical Models (MGMs)[23–25] to describe the associations among the 12 ICP-directed therapies and ICP. An MGM represents each variable as a node and each conditional dependence between two variables as an edge, and it handles the mix of continuous and binary variables in these data. Edge width is proportional to the absolute value of the estimated association parameter, so wider edges indicate stronger conditional associations[25]. We reviewed the resulting graph visually and used it only descriptively to confirm that the selected

therapies were conditionally associated with ICP and with one another. It was not used to select variables, and mortality was not included as a node.

All 12 components were entered into an exploratory multivariable logistic regression model of mortality. Each component was then assigned an integer weight approximating the relative intensity of the therapy, informed by, but not calculated from, its odds ratio. Where an odds ratio was uninformative, for instance if a therapy was near-universal in the cohort (external ventricular drain, analgesia/sedation bolus) or sparsely recorded in the source data (pentobarbital infusion), the weight instead reflected the therapy's position in the guideline tier structure, such that second-tier therapies could not score below first-tier therapies. ePILOT is therefore an ordinal scale of treatment intensity, not a mortality prediction model. Its weights should not be read as risk coefficients, and its association with outcome is evaluated separately. Components are summed, except for $PaCO_2$ and serum sodium, where each contributes only its highest band (at most 3 and 7 points, respectively), so the two bands of either are never counted together. The maximum hourly score is 24 (**Table 3**). For each patient, we calculated the median hourly ePILOT over the first 7 days, and an aggregate ePILOT that counts each therapy once if it was used at any point in those 7 days.

Data were summarized with descriptive statistics, and temporal analyses were performed for the first 7 days of hospitalization. Hourly time series regressions were constructed using Locally Estimated Scatterplot Smoothing (LOESS) for less than 1,000 observations and Generalized Additive Models (GAM) with integrated smoothness estimation for more than 1,000 observations[26–28]. ePILOT scores were also compared against ICP where hourly observations were averaged within each patient and ICP bin such that every patient contributed a single value per bin, and groups were compared using the Wilcoxon rank-sum test. This analysis was

restricted to ICP < 60 mmHg, above which only 12 of 135 patients and 1.8% of monitored hours contributed data. The patient-level association between mean ICP and median hourly ePILOT over the first 7 days was summarized by Spearman's ρ. Optimal cut-points for dichotomizing mortality were computed by a bootstrapping method that maximized Youden's index[29,30].

All models of in-hospital mortality were fitted at the patient level, with each patient contributing a single observation (n = 135, 22 events) where hourly measurements were summarized per patient before analysis. Discrimination was quantified by the area under the receiver operating characteristic curve (AUROC) and the area under the precision-recall curve (AUPRC), with 95% confidence intervals from 2000 percentile bootstrap replicates resampling patients. Calibration was assessed out of sample using 5-fold cross-validation repeated 20 times and is reported as the calibration slope and intercept together with observed and predicted risk by quintile, with the Hosmer–Lemeshow test reported for completeness. Internal validation used bootstrap optimism correction (500 resamples, each refitting the model) alongside the same repeated cross-validation. The incremental value of ePILOT over ICP and admission GCS was assessed by the change in AUROC, a likelihood-ratio test of the nested models, and category-free net reclassification improvement and integrated discrimination improvement, with both models refitted within every bootstrap replicate. The comparison with admission GCS and ICP was repeated using ICP and ePILOT summarized over the first 24 hours of monitoring only. Confidence intervals were obtained by resampling patients rather than patient-hours. Exposure to each component was summarized as the number of patients scored for it in at least one hour. In a sensitivity analysis, the association of median hourly ePILOT with mortality was re-estimated with adjustment for DC and with a DC interaction term, and discrimination and the Youden cut-point were re-derived in patients who never underwent DC.

As an exploratory analysis, we examined the directional structure among ICP, ePILOT, admission GCS, and mortality by constructing a Directed Acyclic Graph (DAG) with the PC algorithm (pcalg package in R)[31–36], with the resulting graph shown in **Supplemental Figure 1**.

All results are presented as frequency (%), mean ± standard deviation or median [interquartile range], as appropriate. All computations and analyses were performed in R version 4.1.0 (www.r-project.org and RStudio, Boston, MA) using cowplot, cutpointr, ggpubr, graph, mgm, pcalg, qgraph, Rgraphviz, scales and tidyverse, and in Python 3.13 using pandas, NumPy, SciPy, statsmodels, scikit-learn, causal-learn, Matplotlib and seaborn. R/Python analysis code was written with the assistance of Claude (Anthropic), a large language model, which was also used to edit manuscript text and to check reference metadata. The authors reviewed and verified all code, outputs and text and take full responsibility for the content.

RESULTS

Baseline demographic and clinical data are shown in **Table 1**. From January 2009 to April 2019, 135 ICP-monitored TBI patients were analyzed. Patients had a median age of 4.8 [1.9–12.7] years; 65% were male and 75% identified as Caucasian. The median admission GCS score was 4 [3–7], and the median admission PELOD-2 score was 13 [10–18]. The median hospital length of stay (HLOS) was 17.0 [11.0–29.0] days and 16% of patients died prior to hospital discharge. Additional patient characteristics stratified by mortality are included in **Table 2**. Patients with in-hospital mortality had higher admission PELOD-2 scores (20 [18–23] vs. 12 [10–16]; $p < 0.001$), had a greater proportion of monitored hours with ICP > 20 mmHg (25 [1–74] vs. 1 [0–4]%; $p = 0.001$), had higher peak ICP (34 [24–72] vs. 26 [20–30] mmHg; $p = 0.008$) and higher mean ICP (26 (± 24) vs. 12 (± 7) mmHg; $p < 0.001$). There were no significant

differences in age, sex, race, or admission GCS score between the two groups. DC was performed in 46 patients (34%), in 4 (18%) who died and 42 (37%) who survived ($p = 0.14$).

Figure 1 is a smoothed hourly time-series plot of ICP values stratified by mortality. Over the first 7 days, ICP values tended to be higher in patients who died. **Figure 2** depicts the MGM that illustrates the different therapies used to manage intracranial hypertension, and their respective relationships. The strongest conditional associations linked therapies that are related by either construction or protocol, thereby supporting the face validity of the MGM. Specifically, there were strong associations between the presence of an EVD and CSF drainage, ranges of hypernatremia (sodium 150–160 mEq/L and > 160 mEq/L), and ranges of $PaCO_2$ (30–35 mmHg and < 30 mmHg). The ePILOT scale is shown in **Table 3**. ePILOT scores range from 0 to 24 where higher scores signify an increased intensity of medical therapeutic support. All components are re-scored hourly except $PaCO_2$, serum sodium and temperature, which can be carried forward up to 4 hours.

ePILOT scores were higher over time in patients who died (**Figure 3**). ePILOT scores increased progressively with ICP up to 60 mmHg (**Figure 4A**), and patients with higher mean ICP had higher median hourly ePILOT scores (Spearman's $\rho = 0.32$, $p < 0.001$). Non-survivors had significantly higher ePILOT scores than survivors at ICP 0–19 mmHg ($p = 0.004$) and 20–39 mmHg ($p < 0.001$), but not at 40–59 mmHg (**Figure 4B**).

The median hourly ePILOT scores over the first 7 days were significantly different between non-survivors and survivors (7.0 [5.0–10.4] vs. 3.5 [3.0–6.5]; $p \leq 0.01$, **Figure 5A**), as were the aggregate ePILOT scores for the first 7 days (15.0 [11.3–19.3] vs. 12.0 [10.0–16.0]; $p \leq 0.05$, **Figure 5B**).

Peak hourly ePILOT scores discriminated in-hospital mortality with an AUROC of 0.74 (95% CI 0.61–0.85, **Supplemental Table 2**), numerically higher than AUROCs of 0.68–0.72 for ICP summaries and 0.61 for admission GCS. Median hourly ePILOT had an AUROC of 0.72 (0.59–0.83) and aggregate 0.66 (0.53–0.78). By AUPRC, however, ICP summaries (0.48–0.51) exceeded the hourly ePILOT summaries (0.36–0.39), against a no-skill baseline of 0.16. Cross-validated calibration slopes were 0.86, 0.91 and 0.80 for the median hourly, peak hourly and aggregate scores, with intercepts of −0.01 to 0.00 and Hosmer–Lemeshow $p \geq 0.48$ (**Supplemental Table 3**). Adding median hourly ePILOT to mean ICP increased the AUROC by 0.098 (95% CI −0.006 to +0.236; likelihood-ratio p = 0.021, continuous net reclassification improvement +0.51, integrated discrimination improvement +0.032), and by 0.077 (−0.005 to +0.185; p = 0.031) when added to admission GCS and ICP (**Supplemental Table 4**). Using only the first 24 hours of monitoring, median hourly ePILOT had an AUROC of 0.75 (0.61–0.87), against 0.65 for mean ICP, and adding it to admission GCS and mean ICP increased the AUROC by 0.080 (−0.010 to +0.204; likelihood-ratio $p = 0.002$). Bootstrap optimism was ≤ 0.005 in absolute AUROC, and cross-validated estimates were 0.69, 0.72 and 0.64 (**Supplemental Table 5**). Using Youden's method, the optimal hourly ePILOT score threshold for dichotomizing mortality was 5 in the first 7 days of hospitalization (**Figure 6**), derived from pooled hourly observations and confirmed to be 5 when re-derived at the patient level (95% CI 4–8; sensitivity 0.77, specificity 0.61, **Supplemental Table 6**).

Hypothermia and serum sodium > 160 mEq/L, the two most heavily weighted components, were each recorded in 64% of patients who died, against 17% and 23% of survivors, whereas EVD presence, CSF drainage, analgesia/sedation boluses and NMB were more common among survivors, who were monitored for longer (**Supplemental Table 7**).

Adjusting for DC did not change the association between median hourly ePILOT and mortality (odds ratio per point 1.26 [1.09–1.44] unadjusted and 1.26 [1.10–1.45] adjusted; interaction p = 0.88). Among the 89 patients who never underwent DC (18 deaths), the AUROC was 0.74 (0.60–0.86) and the Youden cut-point remained 5 (**Supplemental Table 8**).

In an exploratory DAG analysis (**Supplemental Figure 1**), ICP and ePILOT were connected, consistent with a reactive treatment response to rising ICP. Admission GCS, ICP, and ePILOT were each also associated with mortality.

DISCUSSION

In this single-center study, we developed ePILOT, which scores the cumulative intensity of ICP-directed medical therapy every hour from routinely stored EMR data rather than by daily manual chart abstraction. Therapeutic intensity tracked ICP: ePILOT rose progressively with ICP, peaked after ICP did, and, at the same ICP below 40 mmHg, was higher in children who died, consistent with a reactive treatment response to worsening intracranial physiology. The hourly ePILOT can be temporally co-registered with hourly ICP to obtain a more granular determination of the burden of intracranial hypertension in children with TBI vs. measurement of ICP alone. Both median hourly ePILOT and aggregate ePILOT scores were associated with in-hospital mortality in children with severe TBI.

Although used only descriptively, the MGM showed face validity: its strongest associations linked components related by construction or protocol, and the moderate associations of hypothermia (32–34 °C) with analgesia/sedation boluses and NMB are consistent with clinical practice. Shivering during induced hypothermia commonly drives additional

sedation and, when that is insufficient, NMB. To our knowledge, this is the first use of MGMs to describe the inter-relationships among ICP-directed therapies.

We also examined the directional structure among these variables (**Supplemental Figure 1**). Because constraint-based structure learning applied to autocorrelated repeated measures can manufacture apparent directionality, we treat this analysis as hypothesis-generating only, and the no-unmeasured-confounding assumption such inference would require is in any case likely violated by injury severity[37,38]. The reactive-treatment-response interpretation rests instead on the temporal lag between the ICP and ePILOT peaks (**Figure 1**, **Figure 3**). A recent CENTER-TBI analysis in adults found therapeutic intensity, but not ICP alone, to be directly related to outcome[39]. In our cohort both were associated with mortality, but ePILOT had a numerically higher AUROC than any summary of ICP, though not a higher AUPRC, consistent with intensity of therapies carrying intrinsic severity information that ICP alone cannot provide.

Each component's weight was set by the authors, informed by its association with in-hospital mortality and adjusted where an odds ratio was uninformative, such that the scale respects the guideline tier structure. The highest weights fall on second-tier therapies of the 2019 pediatric consensus and guidelines-based algorithm[4]: hypothermia (32–34 °C), hyperventilation ($PaCO_2$ < 30 mmHg) and higher levels of osmotherapy, for which serum sodium serves here as a proxy. Pentobarbital infusion is the exception: its weight was assigned by tier position and carries 2 points.

ePILOT also captures EVD presence, which earns 2 points regardless of CSF drainage (as opposed to an intraparenchymal monitor, which earns no points for presence). This did not inflate the association with mortality – EVDs were more common in survivors, and removing the component did not weaken discrimination (median hourly AUROC 0.75 vs. 0.72). For

comparisons of monitored with unmonitored care, as in the BEST-TRIP trial[40], or across centers with different monitoring practices, the EVD component should be omitted so that only medical therapies are scored.

The hourly and aggregate summaries are complementary: hourly evaluations provide a more dynamic assessment of a patient's needs, while the aggregate score offers an overview of the types of ICP-directed therapies used in the patient's care. The hourly summaries discriminated mortality better than the aggregate, partly because non-survivors were monitored for fewer hours (median 77 vs. 163). Restricted to the first 48 hours, the aggregate score had an AUROC of 0.72, against 0.66 over 7 days. ICP peaked about 43 hours after admission, particularly among non-survivors, while therapy intensity rose steeply between 40 and 55 hours and peaked at about 9 points near 66 hours (**Figure 3**). This lag supports the idea that clinicians react to the patient's worsening intracranial physiology. Measures such as dynamic intracranial compliance might enable more proactive management[41]. As ICP normalizes and therapies are withdrawn, the ePILOT score falls in step, quantifying the weaning process.

At the bedside, the Youden-optimal score of 5 most often corresponds to an EVD with active drainage and one or two first-tier therapies, or an EVD with a single higher-weighted component such as sodium 150–160 mEq/L or $PaCO_2$ < 30 mmHg. A median hourly score of at least 5 was reached by 77% of non-survivors and 39% of survivors. The cut-point separates the two outcome distributions and so falls between the group medians in **Figure 5**; at 16% mortality, a score of 5 or more has a positive predictive value of only about 28%, so it should not be read as a bedside threshold or transported to other centers. ePILOT added information to both ICP and admission GCS, consistent in direction across the three incremental-value measures, although the bootstrap interval for the change in AUROC included zero at 22 events. Since these are single-

predictor summaries, the near-ideal calibration and small optimism reported here address the score's use as a measure and not the derivation of its 12 component weights.

Figure 4 is, in effect, a single center's treatment-response curve. At the same ICP below 40 mmHg, non-survivors received more therapy than survivors (**Figure 4B**), and therefore ePILOT distinguishes an ICP controlled with intensive therapy from the same ICP without it. Generated at other centers, or before and after a protocol change, the curve could be used to compare ICP-targeting strategies.

ePILOT is complementary to automated measures of the physiological insult itself, such as the KidsBrainIT visualization of pediatric ICP and time burden: that work quantifies the burden of intracranial hypertension, whereas ePILOT quantifies the therapeutic response, and the two may be most informative together[42]. Alongside emerging work on TBI phenotypes and biomarkers, a standardized, automated therapy-intensity measure could help assess and compare therapeutic strategies[43] and provide a platform for auditing institutional or provider-level practice across neurocritical care centers.

This study has several limitations. The component weights were set by the authors rather than by a formal consensus process and were likely influenced by institutional practice patterns, although the higher-weighted components correspond to the higher guideline tiers[4]. ePILOT was also not compared with concurrently scored PILOT or TIL, as neither was scored in this cohort, so its agreement with the established scales remains to be tested. Additionally, DI can occur after severe TBI and raise serum sodium independently of therapy. We nonetheless scored serum sodium as it may reflect how much hyperosmolar therapy a patient actually received, whereas binarized scoring for hypertonic saline infusion ignores dose. Future studies should test ePILOT after excluding patients with DI. More generally, ePILOT scores the physiological state that

therapy produces as well as the therapy itself. A $PaCO_2$ below 30 mmHg was recorded in 117 of 135 patients (87%), likely capturing overshoot of $PaCO_2$ targets. Likewise, an EVD scores points whether or not CSF is drained, as discussed above. ePILOT also excludes surgical interventions such as hematoma evacuation or DC. Since DC controls ICP[44], it can lower subsequent medical therapy and thus the ePILOT score. Adjusting for DC did not change the association between ePILOT and mortality, but ePILOT scores before and after DC were not delineated. Surgery was deliberately excluded from ePILOT, as its indications and timing rest with the neurosurgical team. A future aim is to develop cut-offs that signal objectively to consulting teams when a patient's intracranial hypertension is becoming refractory and surgical intervention may be needed. Additionally, in-hospital mortality is a coarse endpoint, and with 22 events (16%) our estimates are imprecise. Future studies could focus on functional outcomes by assessing global outcome with the pediatric version of the Glasgow Outcome Scale–Extended (GOS-E Peds)[45] or a combination of measures spanning adaptive and daily-living skills and health-related quality of life[46]. Finally, the data span 2009–2019, and evolving TBI management recommendations and practice may affect the score over time; prospective evaluation and multicenter external validation are needed.

CONCLUSIONS

ePILOT scores the intensity of ICP-directed medical therapy in children with TBI every hour, automatically, from routinely stored EMR data rather than by manual chart abstraction. Therapeutic intensity rose with ICP, followed it in time, and was higher in children who died, including at the same level of ICP. ePILOT is an intended measure of treatment intensity rather than a prognostic model and its associations with mortality support its construct validity. Future studies should compare ePILOT directly with manually scored PILOT and TIL, validate it

externally at centers with different monitoring practices, relate it to functional outcome, and test it as the treatment-side input, alongside continuous measures of intracranial compliance, ICP waveform morphology and electroencephalography, for automated decision support for ICP-directed therapy. Additional research should also determine the most efficient way to integrate ePILOT seamlessly into the existing clinical workflow.

END MATTER

Data availability

The data supporting this study contain hourly clinical service dates and therefore constitute a limited data set that cannot be shared publicly. De-identified data may be made available to qualified investigators on reasonable request to the corresponding author, subject to a data use agreement and institutional review board approval.

Code availability

Analysis code, including the complete ePILOT scoring implementation, is available from the corresponding author on reasonable request.

Conflicts of interest

The authors declare no competing interests.

Protocol and registration

No study protocol was prepared in advance, and the study was not registered.

Patient and public involvement

There was no patient or public involvement in the design, conduct, reporting or dissemination of this study.

REFERENCES

1. Faul M, Xu L, Wald MM, Coronado VG. Traumatic Brain Injury in the United States: Emergency Department Visits, Hospitalizations and Deaths 2002–2006. Atlanta, GA: Centers for Disease Control and Prevention, National Center for Injury Prevention and Control; 2010.

2. Chesnut R, Videtta W, Vespa P, Le Roux P. Intracranial Pressure Monitoring: Fundamental Considerations and Rationale for Monitoring. Neurocrit Care 2014;21(Suppl 2):S64–84. https://doi.org/10.1007/s12028-014-0048-y. 10.1007/s12028-014-0048-y.

3. Okonkwo DO, Shutter LA, Moore C, et al. Brain Oxygen Optimization in Severe Traumatic Brain Injury Phase-II: A Phase II Randomized Trial. Crit Care Med 2017;45(11):1907–14. https://doi.org/10.1097/CCM.0000000000002619. 10.1097/CCM.0000000000002619.

4. Kochanek PM, Tasker RC, Bell MJ, et al. Management of Pediatric Severe Traumatic Brain Injury: 2019 Consensus and Guidelines-Based Algorithm for First and Second Tier Therapies. Pediatr Crit Care Med 2019;20(3):269–79. https://doi.org/10.1097/PCC.0000000000001737. 10.1097/PCC.0000000000001737.

5. Hawryluk GWJ, Aguilera S, Buki A, et al. A management algorithm for patients with intracranial pressure monitoring: the Seattle International Severe Traumatic Brain Injury Consensus Conference (SIBICC). Intensive Care Med 2019;45(12):1783–94. https://doi.org/10.1007/s00134-019-05805-9. 10.1007/s00134-019-05805-9.

6. Agrawal S, Bögli SY, O'Donnell R, et al. Intracranial Pressure Treatment Thresholds in Pediatric Traumatic Brain Injury. JAMA Pediatr 2026;180(6):650. https://doi.org/10.1001/jamapediatrics.2026.0197. 10.1001/jamapediatrics.2026.0197.

7. Simon DW, Shaik MA, Kochanek PM. Challenging 20 mm Hg as the ICP Threshold in Pediatric TBI. JAMA Pediatr 2026;180(6):607. https://doi.org/10.1001/jamapediatrics.2026.0820. 10.1001/jamapediatrics.2026.0820.

8. Steyerberg EW, Mushkudiani N, Perel P, et al. Predicting Outcome after Traumatic Brain Injury: Development and International Validation of Prognostic Scores Based on Admission Characteristics. PLoS Med 2008;5(8):e165. https://doi.org/10.1371/journal.pmed.0050165. 10.1371/journal.pmed.0050165.

9. Maset AL, Marmarou A, Ward JD, et al. Pressure-volume index in head injury. J Neurosurg 1987;67(6):832–40. https://doi.org/10.3171/jns.1987.67.6.0832. 10.3171/jns.1987.67.6.0832.

10. Shore PM, Hand LL, Roy L, Trivedi P, Kochanek PM, Adelson PD. Reliability and validity of the Pediatric Intensity Level of Therapy (PILOT) scale: A measure of the use of intracranial pressure–directed therapies. Crit Care Med 2006;34(7):1981–7. https://doi.org/10.1097/01.CCM.0000220765.22184.ED. 10.1097/01.CCM.0000220765.22184.ED.

11. Zuercher P, Groen JL, Aries MJH, et al. Reliability and Validity of the Therapy Intensity Level Scale: Analysis of Clinimetric Properties of a Novel Approach to Assess Management of Intracranial Pressure in Traumatic Brain Injury. J Neurotrauma 2016;33(19):1768–74. https://doi.org/10.1089/neu.2015.4266. 10.1089/neu.2015.4266.

12. Maas AIR, Harrison-Felix CL, Menon D, et al. Standardizing Data Collection in Traumatic Brain Injury. J Neurotrauma 2011;28(2):177–87. https://doi.org/10.1089/neu.2010.1617. 10.1089/neu.2010.1617.

13. Bhattacharyay S, Beqiri E, Zuercher P, et al. Therapy Intensity Level Scale for Traumatic Brain Injury: Clinimetric Assessment on Neuro-Monitored Patients Across 52 European Intensive Care Units. J Neurotrauma 2024;41(7–8):887–909. https://doi.org/10.1089/neu.2023.0377. 10.1089/neu.2023.0377.

14. Zeiler FA, Ercole A, Beqiri E, et al. Cerebrovascular reactivity is not associated with therapeutic intensity in adult traumatic brain injury: a CENTER-TBI analysis. Acta Neurochir (Wien) 2019;161(9):1955–64. https://doi.org/10.1007/s00701-019-03980-8. 10.1007/s00701-019-03980-8.

15. von Elm E, Egger M, Altman DG, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. Strengthening the reporting of observational studies in epidemiology (STROBE) statement: Guidelines for reporting observational studies. BMJ 2007;335(7624):806–8. https://doi.org/10.1136/BMJ.39386.490150.94. 10.1136/BMJ.39386.490150.94.

16. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ 2024;385:e078378. https://doi.org/10.1136/bmj-2023-078378. 10.1136/bmj-2023-078378.

17. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ 2015;350:g7594. https://doi.org/10.1136/bmj.g7594. 10.1136/bmj.g7594.

18. Hripcsak G, Duke JD, Shah NH, et al. Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. Stud Health Technol Inform 2015;216:574–8.

19. Adelson P, Bratton S, Carney N, et al. Guidelines for the acute medical management of severe traumatic brain injury in infants, children, and adolescents. Pediatr Crit Care Med 2003;4(3 Suppl):S1–75.

20. Kochanek P, Carney N, Adelson P, et al. Guidelines for the acute medical management of severe traumatic brain injury in infants, children, and adolescents–second edition. Pediatr Crit Care Med 2012;13 Suppl 1:S1–82. https://doi.org/10.1097/PCC.0B013E31823F435C. 10.1097/PCC.0B013E31823F435C.

21. Kochanek PM, Tasker RC, Carney N, et al. Guidelines for the Management of Pediatric Severe Traumatic Brain Injury, Third Edition: Update of the Brain Trauma Foundation

Guidelines. Pediatr Crit Care Med 2019;20(3S Suppl 1):S1–82. https://doi.org/10.1097/PCC.0000000000001735. 10.1097/PCC.0000000000001735.

22. Kochanek PM, Adelson PD, Rosario BL, et al. Comparison of Intracranial Pressure Measurements Before and After Hypertonic Saline or Mannitol Treatment in Children With Severe Traumatic Brain Injury. JAMA Netw Open 2022;5(3):e220891. https://doi.org/10.1001/jamanetworkopen.2022.0891. 10.1001/jamanetworkopen.2022.0891.

23. Sedgewick AJ, Shi I, Donovan RM, Benos PV. Learning mixed graphical models with separate sparsity parameters and stability-based model selection. BMC Bioinformatics 2016;17(S5):S175. https://doi.org/10.1186/s12859-016-1039-0. 10.1186/s12859-016-1039-0.

24. Raghu VK, Horvat CM, Kochanek PM, et al. Neurological Complications Acquired During Pediatric Critical Illness: Exploratory "Mixed Graphical Modeling" Analysis Using Serum Biomarker Levels. Pediatr Crit Care Med 2021;22(10):906–14. https://doi.org/10.1097/PCC.0000000000002776. 10.1097/PCC.0000000000002776.

25. Haslbeck JMB, Waldorp LJ. mgm: Estimating Time-Varying Mixed Graphical Models in High-Dimensional Data. J Stat Softw 2020;93:1–46. https://doi.org/10.18637/JSS.V093.I08. 10.18637/JSS.V093.I08.

26. SAS Institute Inc. SAS/STAT 14.1 User's Guide: The LOESS Procedure. Cary, NC: SAS Institute Inc.; 2015.

27. Wood SN. Generalized Additive Models: An Introduction with R. Chapman and Hall/CRC; 2017.

28. Cleveland WS, Grosse E, Shyu WM. Local regression models. In: Chambers JM, Hastie TJ, editors. Statistical Models in S. Pacific Grove, CA: Wadsworth & Brooks/Cole; 1992. p. 309–76.

29. Fluss R, Faraggi D, Reiser B. Estimation of the Youden Index and its associated cutoff point. Biom J 2005;47(4):458–72. https://doi.org/10.1002/bimj.200410135. 10.1002/bimj.200410135.

30. Perkins NJ, Schisterman EF. The Youden index and the optimal cut-point corrected for measurement error. Biom J 2005;47(4):428–41. https://doi.org/10.1002/bimj.200410133. 10.1002/bimj.200410133.

31. Pearl J. Causal diagrams for empirical research. Biometrika 1995;82(4):669–88. https://doi.org/10.1093/biomet/82.4.669. 10.1093/biomet/82.4.669.

32. Williams TC, Bach CC, Matthiesen NB, Henriksen TB, Gagliardi L. Directed acyclic graphs: a tool for causal studies in paediatrics. Pediatr Res 2018;84(4):487–93. https://doi.org/10.1038/s41390-018-0071-3. 10.1038/s41390-018-0071-3.

33. Glymour M, Greenland S. Causal Diagrams. Modern Epidemiology. Lippincott Williams & Wilkins; 2008.

34. Kalisch M, Mächler M, Colombo D, Maathuis MH, Bühlmann P. Causal inference using graphical models with the R package pcalg. J Stat Softw 2012;47(11):1–26. https://doi.org/10.18637/jss.v047.i11. 10.18637/jss.v047.i11.

35. Kalisch M, Bühlmann P. Estimating high-dimensional directed acyclic graphs with the PC-algorithm. J Mach Learn Res 2007;8:613–36.

36. Spirtes Peter, Glymour CN, Scheines Richard. Causation, prediction, and search. MIT Press; 2000.

37. Hernán MA. The C-Word: Scientific Euphemisms Do Not Improve Causal Inference From Observational Data. Am J Public Health 2018;108(5):616–9. https://doi.org/10.2105/AJPH.2018.304337. 10.2105/AJPH.2018.304337.

38. Manley GT, Lingsma HF, Maas AIR. ADAPTing to a New Era of Comparative Effectiveness Research in Traumatic Brain Injury—Generating Evidence From Observational Data. JAMA Netw Open 2022;5(3):e220899. https://doi.org/10.1001/jamanetworkopen.2022.0899. 10.1001/jamanetworkopen.2022.0899.

39. Åkerlund CAI, Holst A, Beqiri E, et al. Toward Causal Relationships of Intracranial Pressure, Cerebrovascular Reactivity, Interventions, and Outcome in Severe Traumatic Brain Injury: A CENTER-TBI High-Resolution Substudy. Neurocrit Care May 29, 2026 [Online ahead of print]. https://doi.org/10.1007/s12028-026-02552-w. 10.1007/s12028-026-02552-w.

40. Chesnut RM, Temkin N, Carney N, et al. A Trial of Intracranial-Pressure Monitoring in Traumatic Brain Injury. N Engl J Med 2012;367(26):2471–81. https://doi.org/10.1056/NEJMoa1207363. 10.1056/NEJMoa1207363.

41. Wolf MS, Rakkar J, Horvat CM, et al. Assessment of Dynamic Intracranial Compliance in Children with Severe Traumatic Brain Injury: Proof-of-Concept. Neurocrit Care 2021;34(1):209–17. https://doi.org/10.1007/s12028-020-01004-3. 10.1007/s12028-020-01004-3.

42. Kempen B, Depreitere B, Piper I, et al. Visualization of the Intracranial Pressure and Time Burden in Childhood Brain Trauma: What We Have Learned One Decade on With KidsBrainIT. J Neurotrauma 2024;41(13–14):e1651–9. https://doi.org/10.1089/neu.2023.0254. 10.1089/neu.2023.0254.

43. Jha RM, Elmer J, Zusman BE, et al. Intracranial Pressure Trajectories: A Novel Approach to Informing Severe Traumatic Brain Injury Phenotypes. Crit Care Med 2018;46(11):1792–802. https://doi.org/10.1097/CCM.0000000000003361. 10.1097/CCM.0000000000003361.

44. Young AMH, Kolias AG, Hutchinson PJ. Decompressive craniectomy for traumatic intracranial hypertension: application in children. Childs Nerv Syst 2017;33(10):1745–50. https://doi.org/10.1007/s00381-017-3534-7. 10.1007/s00381-017-3534-7.

45. Beers SR, Wisniewski SR, Garcia-Filion P, et al. Validity of a Pediatric Version of the Glasgow Outcome Scale–Extended. J Neurotrauma 2012;29(6):1126–39. https://doi.org/10.1089/neu.2011.2272. 10.1089/neu.2011.2272.

46. McCauley SR, Wilde EA, Anderson VA, et al. Recommendations for the Use of Common Outcome measures in pediatric traumatic brain injury research. J Neurotrauma 2012;29(4):678–705. https://doi.org/10.1089/neu.2011.1838. 10.1089/neu.2011.1838.

FIGURE CAPTIONS

Figure 1. Temporal trends in ICP in monitored TBI patients over the first 7 days of hospitalization. Gray shading associated with the smoothed-line graph represents the 95% confidence interval (CI). ICP, intracranial pressure; TBI, traumatic brain injury.

Figure 2. Mixed Graphical Model of the 12 ePILOT components (white nodes) and ICP (black node). Each edge is a conditional association between two variables after accounting for all others; edge width is proportional to the absolute strength of the association. The $PaCO_2 < 35$ mmHg node denotes the 30–35 mmHg band. CSF, cerebrospinal fluid; EVD, external ventricular drain; gtt, continuous infusion; HyperOsm, hyperosmolar; ICP, intracranial pressure; NMB, neuromuscular blockade; $PaCO_2$, partial pressure of arterial carbon dioxide.

Figure 3. Temporal trends in ePILOT score in monitored TBI patients over the first 7 days of hospitalization. Gray shading associated with the smoothed-line graph represents the 95% confidence interval (CI). ePILOT, EMR-based Pediatric Intensity Level of Therapy scale; TBI, traumatic brain injury. The dashed vertical line marks the maximum of the smoothed mean score in non-survivors (approximately 66 hours), not the highest individual hourly value.

Figure 4. Comparison of ePILOT score to ICP. Analyses are restricted to ICP < 60 mmHg; higher values were recorded in only 12 of 135 patients and 1.8% of monitored hours. Each monitored hour is assigned to a bin by that hour's ICP, and a patient's hourly ePILOT scores are averaged within each bin, so each patient contributes one value to every bin their ICP entered. (A) Mean ePILOT score across patients against ICP, binned in 2 mmHg increments; shading indicates the 95% confidence interval. (B) Mean (± standard error) ePILOT score against ICP in 20 mmHg bins, stratified by in-hospital mortality. $^{**}p \leq 0.01$, $^{***}p \leq 0.001$ by Wilcoxon rank-

sum test. ePILOT, EMR-based Pediatric Intensity Level of Therapy scale; ICP, intracranial pressure.

Figure 5. Median hourly (A) and aggregate (B) ePILOT scores over the first 7 days of hospitalization in patients who survived versus those who did not survive to hospital discharge. $^{**}p \leq 0.01$, $^{*}p \leq 0.05$. ICP, intracranial pressure; ePILOT, EMR-based Pediatric Intensity Level of Therapy scale; TBI, traumatic brain injury.

Figure 6. Using cut-point analysis, the maximal value for the Youden Index, which reflects the performance of the hourly ePILOT score in dichotomizing mortality, intersects the x-axis at 5 when evaluated over the first 7 days of hospitalization in patients with TBI with ICP monitoring. The gray shaded area represents the 95% CI from bootstrap resampling of hourly observations. Because hourly observations within a patient are correlated, this interval is narrower than one obtained by resampling patients; the cut-point re-derived at the patient level, with patients as the resampling unit, is given in **Supplemental Table 6**. CI, confidence interval; ePILOT, EMR-based Pediatric Intensity Level of Therapy scale.

REPORTING CHECKLISTS

STROBE Statement: checklist of items that should be included in reports of cohort studies

Item wording reproduced from the STROBE Statement (von Elm E, et al. BMJ 2007;335:806–8; www.strobe-statement.org). Locations refer to sections of this manuscript.

Item	Recommendation	Reported in
Title and abstract		
1 (a)	Indicate the study's design with a commonly used term in the title or the abstract	Title; Abstract
1 (b)	Provide in the abstract an informative and balanced summary of what was done and what was found	Abstract
Introduction		
2	Explain the scientific background and rationale for the investigation being reported	Introduction
3	State specific objectives, including any prespecified hypotheses	Introduction, final paragraph
Methods		
4	Present key elements of study design early in the paper	Methods - Study Population and Data Collection
5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	Methods - Study Population and Data Collection
6 (a)	Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	Methods - Study Population
6 (b)	For matched studies, give matching criteria and number of exposed and unexposed	Not applicable
7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	Methods - Patient Management and Data Selection; Table 3
8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	Methods - Study Population; Supplemental Table 1
9	Describe any efforts to address potential sources of bias	Discussion, limitations paragraph
10	Explain how the study size was arrived at	Discussion, limitations paragraph
11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	Methods - Patient Management; Table 3
12 (a)	Describe all statistical methods, including those used to control for confounding	Methods - Statistical Methods
12 (b)	Describe any methods used to examine subgroups and interactions	Methods - Statistical Methods
12 (c)	Explain how missing data were addressed	Methods - Patient Management and Data Selection; Discussion, limitations
12 (d)	If applicable, explain how loss to follow-up was addressed	Not applicable
12 (e)	Describe any sensitivity analyses	Results; Supplemental Table 8
Results		
13* (a)	Report numbers of individuals at each stage of study - eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	Methods - Study Population
13* (b)	Give reasons for non-participation at each stage	Methods - Study Population
13* (c)	Consider use of a flow diagram	Not included; participant counts are given in Methods - Study Population
14* (a)	Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	Tables 1 and 2
14* (b)	Indicate number of participants with missing data for each variable of interest	Tables 1 and 2 (footnote)
14* (c)	Summarise follow-up time (eg, average and total amount)	Table 2
15*	Report numbers of outcome events or summary measures over time	Results; Tables 1 and 2
16 (a)	Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	Results; Supplemental Tables 2 and 4
16 (b)	Report category boundaries when continuous variables were categorized	Table 3; Figure 4
16 (c)	If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	Not applicable

Item	Recommendation	Reported in
17	Report other analyses done - eg analyses of subgroups and interactions, and sensitivity analyses	Results; Figures 2 and 6; Supplemental Figure 1; Supplemental Table 8
Discussion		
18	Summarise key results with reference to study objectives	Discussion, opening paragraph
19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	Discussion, limitations paragraph
20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	Discussion
21	Discuss the generalisability (external validity) of the study results	Discussion, limitations and closing
Other information		
22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	Title page - Support; Details page - Funding

TRIPOD+AI checklist

Item wording reproduced from Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement. BMJ 2024;385:e078378. This is a development study with internal validation; locations refer to sections of this manuscript.

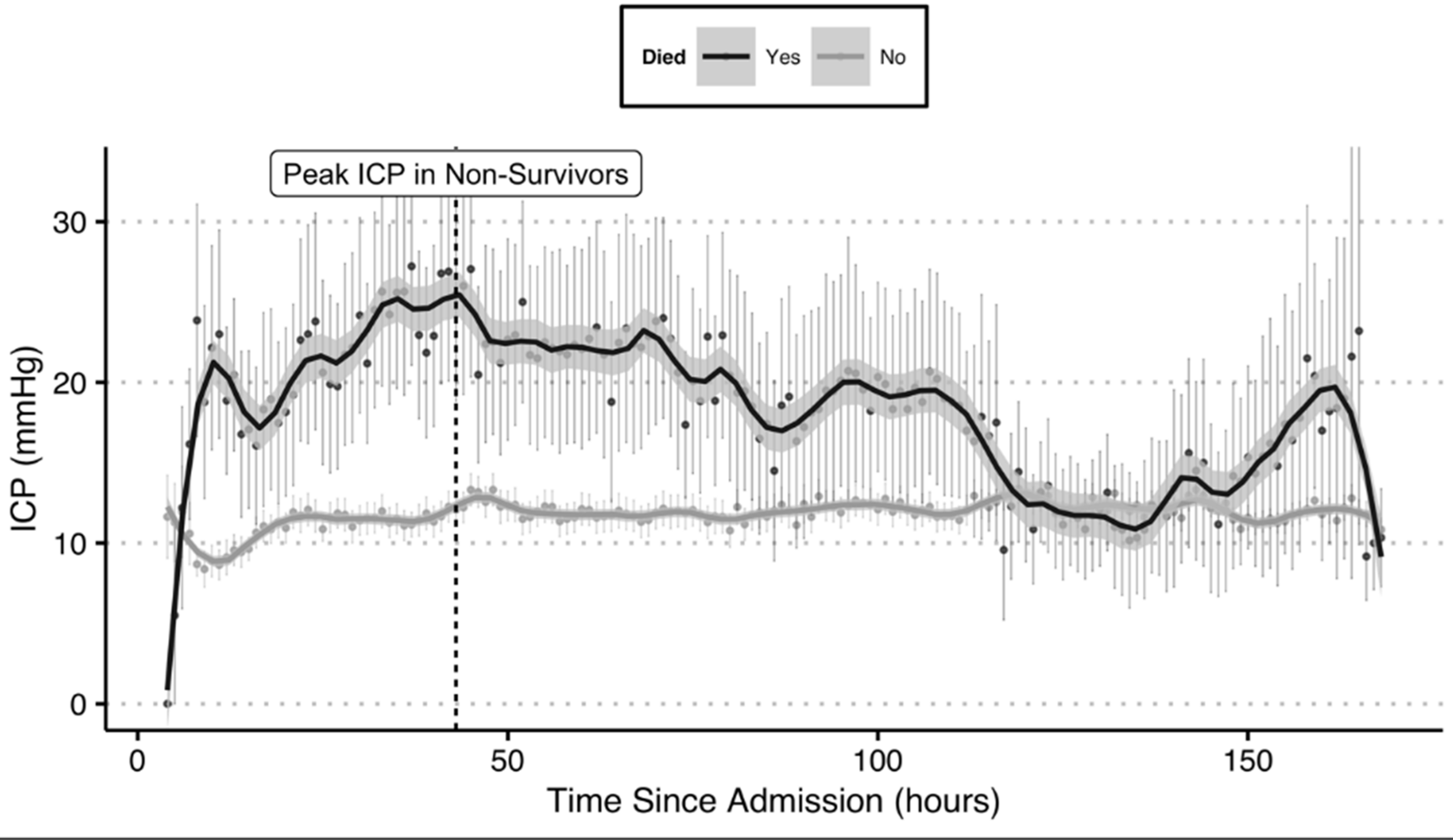
Item	Recommendation	Reported in
Title		
1	Identify the study as developing or evaluating the performance of a multivariable prediction model, the target population, and the outcome to be predicted	Title
Abstract		
2	See TRIPOD+AI for Abstracts checklist	Abstract
Introduction		
3a	Explain the healthcare context (including whether diagnostic or prognostic) and rationale for developing or evaluating the prediction model, with references to existing models	Introduction
3b	Describe the target population and the intended purpose of the prediction model in the context of the care pathway, including its intended users	Introduction; Discussion
3c	Describe any known health inequalities between sociodemographic groups	Discussion, limitations paragraph
4	Specify the study objectives, including whether the study describes the development or validation of a prediction model (or both)	Introduction, final paragraph
Methods		
5a	Describe the sources of data separately for the development and evaluation datasets, the rationale for using these data, and representativeness of the data	Methods - Study Population and Data Collection
5b	Specify the dates of the collected participant data, including start and end of participant accrual; and, if applicable, end of follow-up	Methods - Study Population
6a	Specify key elements of the study setting including the number and location of centres	Methods - Study Population
6b	Describe the eligibility criteria for study participants	Methods - Study Population
6c	Give details of any treatments received, and how they were handled during model development or evaluation, if relevant	Methods - Patient Management and Data Selection
7	Describe any data pre-processing and quality checking, including whether this was similar across relevant sociodemographic groups	Methods - Study Population; Patient Management
8a	Clearly define the outcome that is being predicted and the time horizon, including how and when assessed, the rationale for choosing this outcome, and whether the method of outcome assessment is consistent across sociodemographic groups	Methods - Study Population; Results
8b	If outcome assessment requires subjective interpretation, describe the qualifications and demographic characteristics of the outcome assessors	Not applicable
8c	Report any actions to blind assessment of the outcome to be predicted	Not applicable
9a	Describe the choice of initial predictors and any pre-selection of predictors before model building	Methods - Patient Management; Statistical Methods
9b	Clearly define all predictors, including how and when they were measured (and any actions to blind assessment of predictors)	Methods - Patient Management; Table 3; Supplemental Table 1
9c	If predictor measurement requires subjective interpretation, describe the qualifications and demographic characteristics of the predictor assessors	Not applicable
10	Explain how the study size was arrived at (separately for development and evaluation), and justify that the study size was sufficient to answer the research question	Discussion, limitations paragraph
11	Describe how missing data were handled. Provide reasons for omitting any data	Methods - Patient Management and Data Selection; Discussion, limitations
12a	Describe how the data were used in the analysis, including whether the data were partitioned, considering any sample size requirements	Methods - Statistical Methods
12b	Depending on the type of model, describe how predictors were handled in the analyses (functional form, rescaling, transformation, or any standardisation)	Methods - Patient Management; Table 3
12c	Specify the type of model, rationale, all model building steps, including any hyperparameter tuning, and method for internal validation	Methods - Statistical Methods
12d	Describe if and how any heterogeneity in estimates of model parameter values and model performance was handled and quantified across clusters	Methods - Statistical Methods

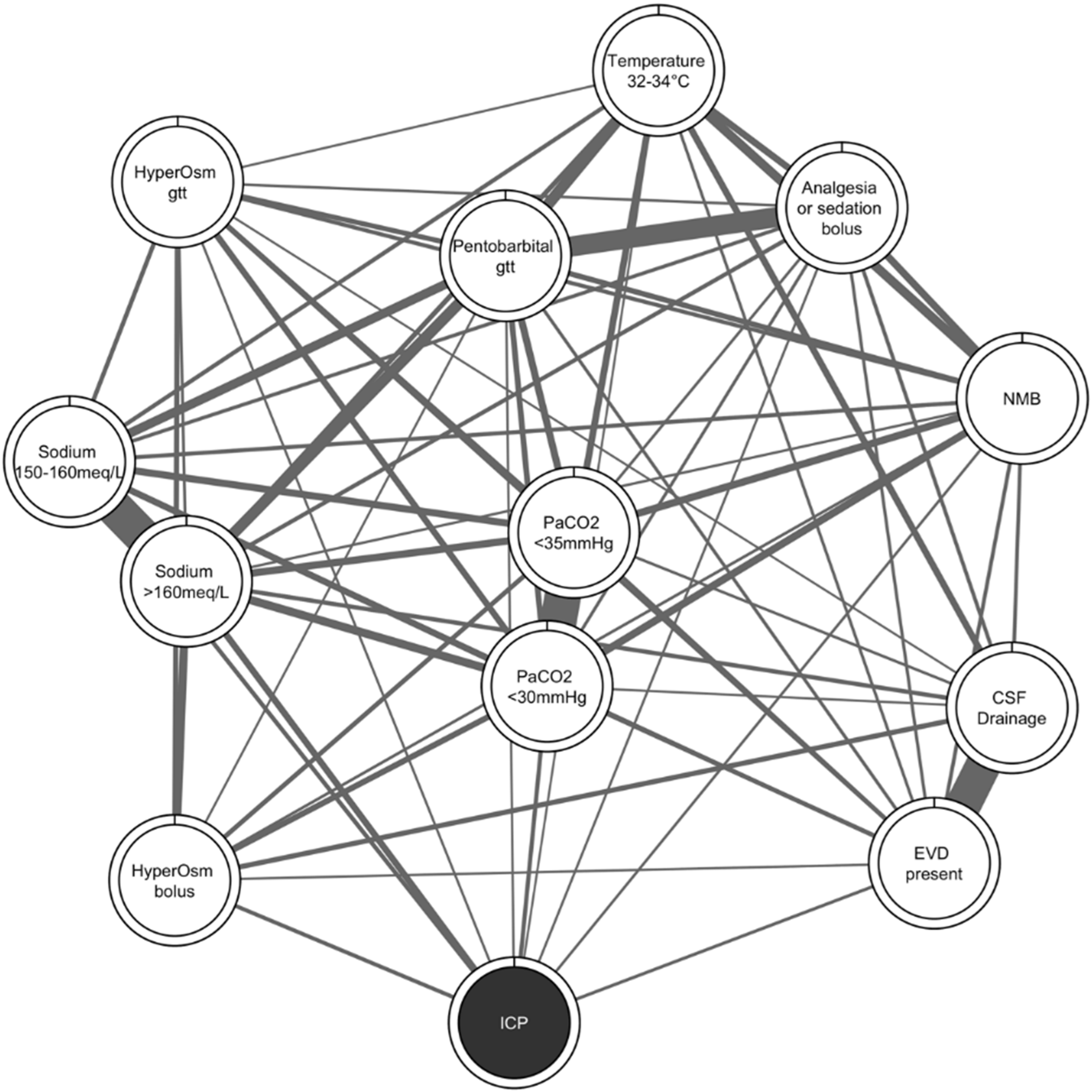
Item	Recommendation	Reported in
12e	Specify all measures and plots used (and their rationale) to evaluate model performance (e.g., discrimination, calibration, clinical utility) and, if relevant, to compare multiple models	Methods - Statistical Methods
12f	Describe any model updating (e.g., recalibration) arising from the model evaluation, either overall or for particular sociodemographic groups or settings	Not applicable
12g	For model evaluation, describe how the model predictions were calculated (e.g., formula, code, object, application programming interface)	Table 3; End Matter - Code availability
13	If class imbalance methods were used, state why and how this was done, and any subsequent methods to recalibrate the model or the model predictions	Discussion, limitations
14	Describe any approaches that were used to address model fairness and their rationale	Discussion, limitations paragraph
15	Specify the output of the prediction model (e.g., probabilities, classification). Provide details and rationale for any classification and how the thresholds were identified	Methods - Statistical Methods; Results; Figure 6
16	Identify any differences between the development and evaluation data in healthcare setting, eligibility criteria, outcome, and predictors	Not applicable
17	Name the institutional research board or ethics committee that approved the study and describe the participant informed consent or the ethics committee waiver of informed consent	Methods - Study Population and Data Collection; Details page - Ethics
Open science		
18a	Give the source of funding and the role of the funders for the present study	Title page - Support; Details page - Funding
18b	Declare any conflicts of interest and financial disclosures for all authors	Details page - Conflicts of interest; End Matter - Conflicts of interest
18c	Indicate where the study protocol can be accessed or state that a protocol was not prepared	End Matter - Protocol and registration
18d	Provide registration information for the study, including register name and registration number, or state that the study was not registered	End Matter - Protocol and registration
18e	Provide details of the availability of the study data	End Matter - Data availability
18f	Provide details of the availability of the analytical code	End Matter - Code availability
Patient and public involvement		
19	Provide details of any patient and public involvement during the design, conduct, reporting, interpretation, or dissemination of the study or state no involvement	End Matter - Patient and public involvement
Results		
20a	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time	Methods - Study Population; Tables 1 and 2
20b	Report the characteristics overall and, where applicable, for each data source or setting, including the key dates, key predictors (including demographics), treatments received, sample size, number of outcome events, follow-up time, and amount of missing data	Tables 1 and 2
20c	For model evaluation, show a comparison with the development data of the distribution of important predictors (demographics, predictors, and outcome)	Not applicable
21	Specify the number of participants and outcome events in each analysis (e.g., for model development, hyperparameter tuning, model evaluation)	Results
22	Provide details of the full prediction model (e.g., formula, code, object, application programming interface) to allow predictions in new individuals and to enable third party evaluation and implementation	Table 3; Supplemental Table 1; End Matter - Code availability
23a	Report model performance estimates with confidence intervals, including for any key subgroups (e.g., sociodemographic). Consider plots to aid presentation	Results; Supplemental Tables 2-6
23b	If examined, report results of any heterogeneity in model performance across clusters	Not applicable (single centre)
24	Report the results from any model updating, including the updated model and subsequent performance	Not applicable
Discussion		

Item	Recommendation	Reported in
25	Give an overall interpretation of the main results, including issues of fairness in the context of the objectives and previous studies	Discussion
26	Discuss any limitations of the study (such as a non-representative sample, sample size, overfitting, missing data) and their effects on any biases, statistical uncertainty, and generalisability	Discussion, limitations
27a	Describe how poor quality or unavailable input data (e.g., predictor values) should be assessed and handled when implementing the prediction model	Methods - Patient Management
27b	Specify whether users will be required to interact in the handling of the input data or use of the model, and what level of expertise is required of users	Discussion
27c	Discuss any next steps for future research, with a specific view to applicability and generalisability of the model	Discussion, closing paragraphs; Conclusions

Time Series Plot of ICP for Monitored Severe TBI Patients

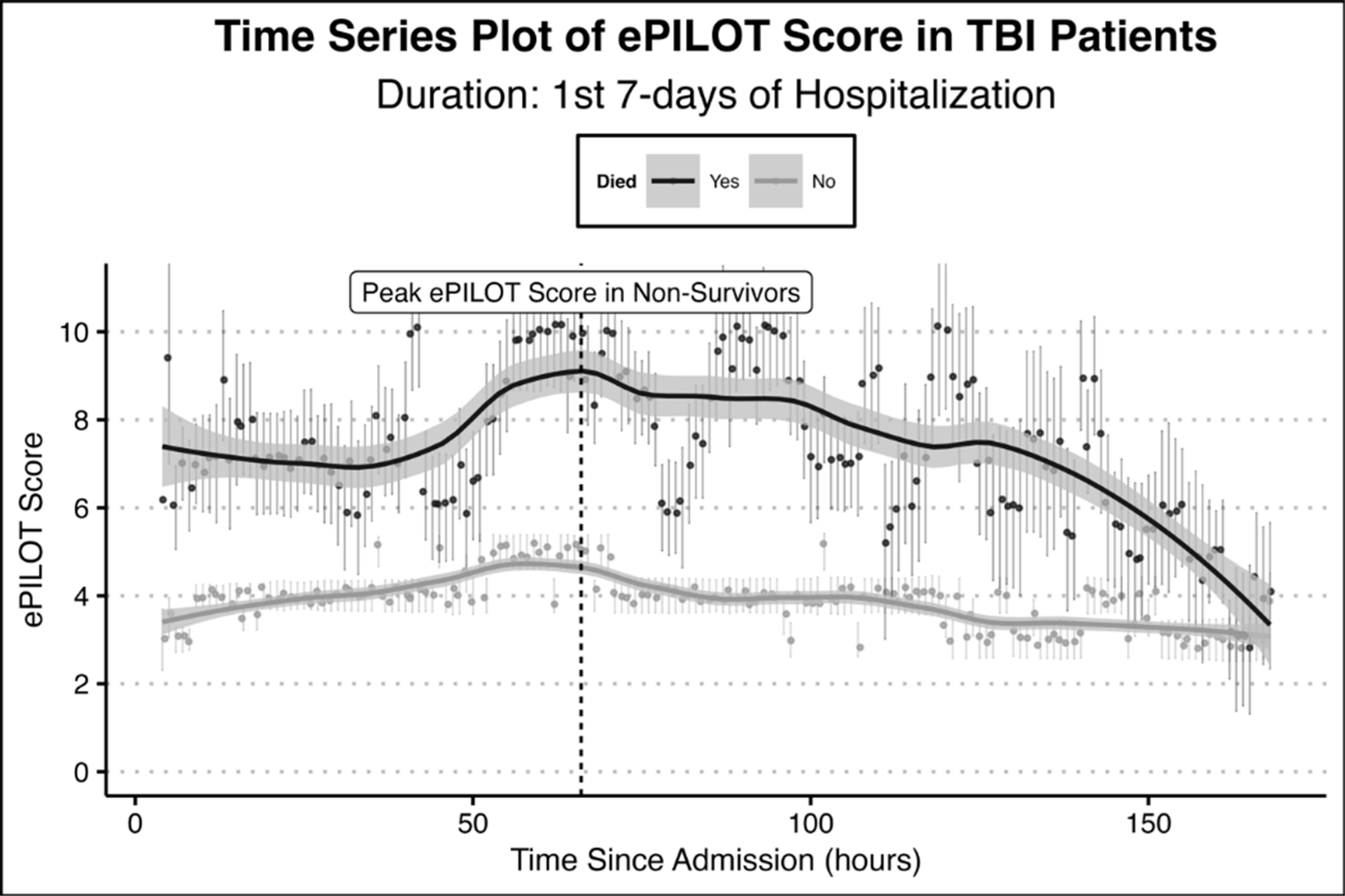
Duration: 1st 7-days of Hospitalization

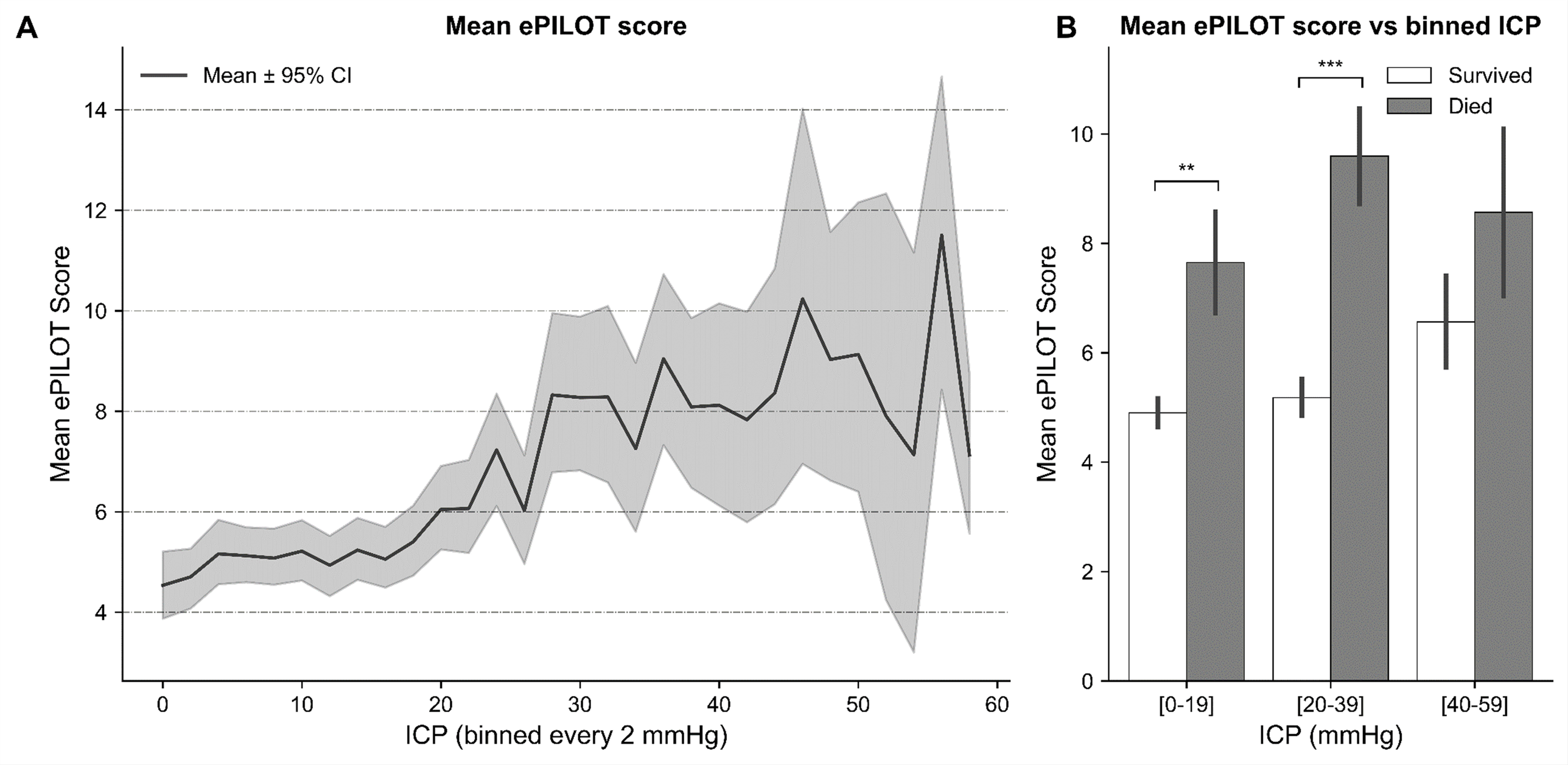


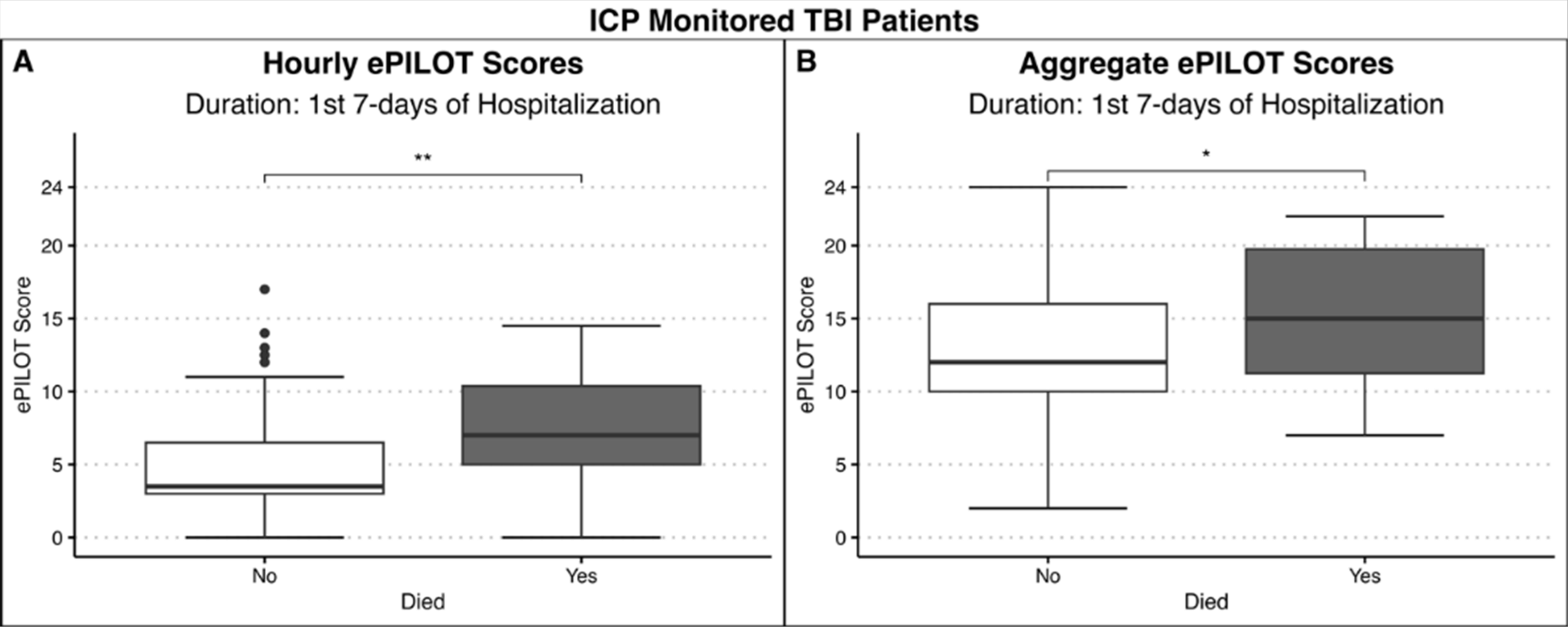

Temperature
32-34°C
HyperOsm
gtt
Pentobarbital
gtt
Analgesia
or sedation
bolus
NMB
Sodium
150-160meq/L
PaCO2
<35mmHg
Sodium
>160meq/L
PaCO2
<30mmHg
CSF
Drainage
HyperOsm
bolus
EVD
present
ICP

Time Series Plot of ePILOT Score in TBI Patients

Duration: 1st 7-days of Hospitalization




A
Mean ePILOT score
Mean ± 95% CI
Mean ePILOT Score
14
12
10
8
6
4
0
10
20
30
40
50
60
ICP (binned every 2 mmHg)
B
Mean ePILOT score vs binned ICP
Survived
Died
**
***
Mean ePILOT Score
10
8
6
4
2
0
[0-19]
[20-39]
[40-59]
ICP (mmHg)


ICP Monitored TBI Patients
A
Hourly ePILOT Scores
Duration: 1st 7-days of Hospitalization
**
ePILOT Score
24
20
15
10
5
0
No
Yes
Died
B
Aggregate ePILOT Scores
Duration: 1st 7-days of Hospitalization
*
ePILOT Score
24
20
15
10
5
0
No
Yes
Died

ePILOT Score Thresholds for Predicting Mortality

Duration: 1st 7-days of Hospitalization

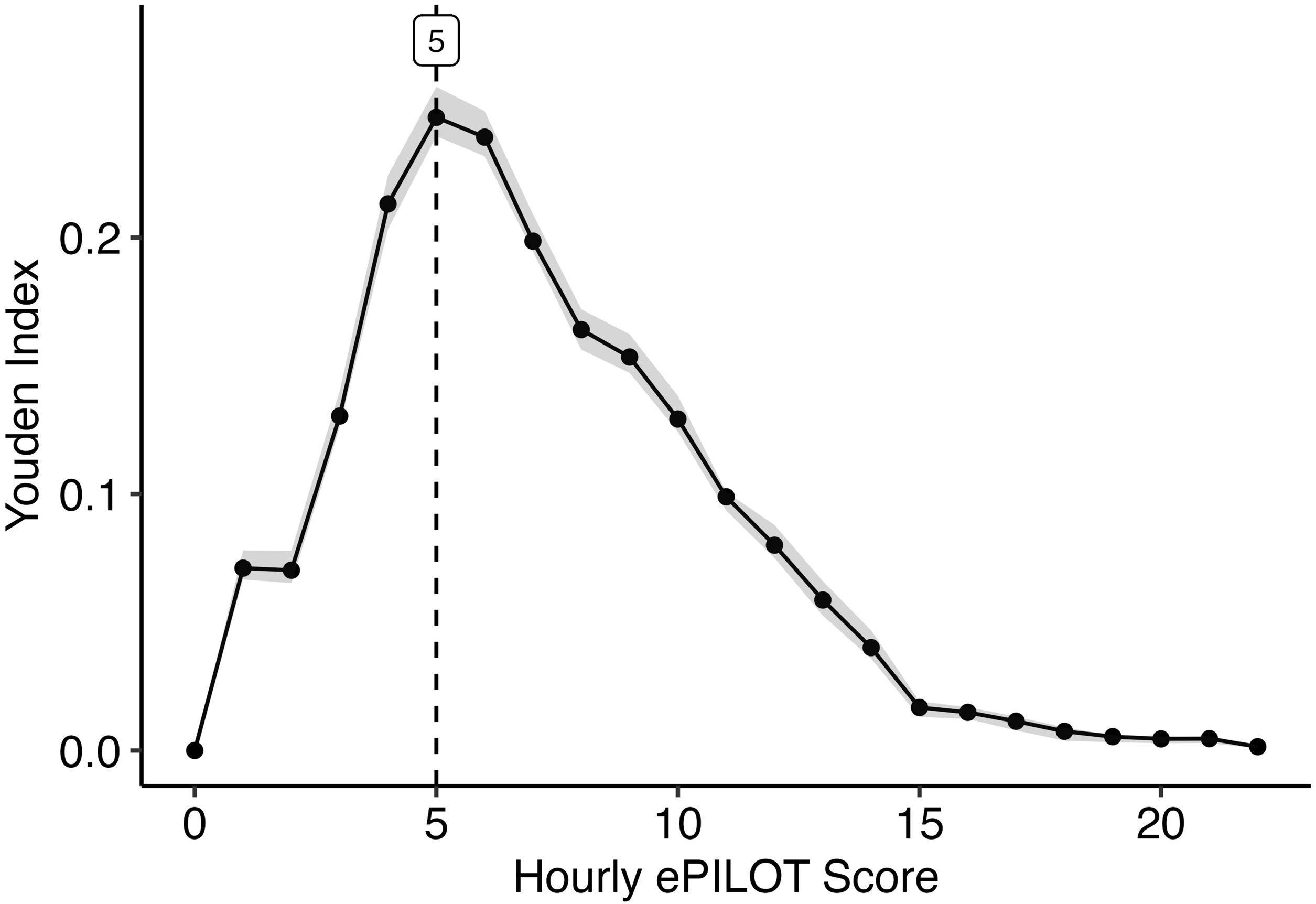


Table 1 Demographics of TBI Patients	
Characteristic	**N = 135**[1]
Age (years)	4.8 [1.9–12.7]
Sex	
Male	88 (65.2%)
Female	47 (34.8%)
Race	
Caucasian	101 (74.8%)
Non-Caucasian	34 (25.2%)
Admission GCS score	4 [3–7]
Admission PELOD-2 score	13 [10–18]
Decompressive craniectomy	46 (34.1%)
Hospital length of stay (days)	17.0 [11.0–29.0]
In-hospital mortality	22 (16.3%)

[1]*Median [IQR]; Frequency (%).*

Table 1. Baseline patient demographics of the study population. PELOD-2 (Pediatric Logistic Organ Dysfunction-2) score is a descriptive scoring system used to evaluate the severity of organ dysfunction. GCS, Glasgow Coma Scale.

Table 2 Mortality Stratified Demographics of TBI patients.				
Characteristic	**Overall, N = 135[1]**	**Yes, N = 22[1]**	**No, N = 113[1]**	**p-value[2]**
Age (years)	4.8 [1.9–12.7]	5.4 [1.4–9.0]	4.8 [1.9–13.2]	0.723
Sex				0.368
Male	88 (65.2%)	12 (54.5%)	76 (67.3%)	
Female	47 (34.8%)	10 (45.5%)	37 (32.7%)	
Race				>0.999
Caucasian	101 (74.8%)	16 (72.7%)	85 (75.2%)	
Non-Caucasian	34 (25.2%)	6 (27.3%)	28 (24.8%)	
Admission GCS score	4 [3–7]	3 [3–5]	4 [3–7]	0.080
Admission PELOD-2 score	13 [10–18]	20 [18–23]	12 [10–16]	<0.001
Monitored with an EVD	117 (86.7%)	15 (68.2%)	102 (90.3%)	0.012
Decompressive craniectomy	46 (34.1%)	4 (18.2%)	42 (37.2%)	0.138
Monitored hours with ICP > 20 mmHg (%)	2 [0–6]	25 [1–74]	1 [0–4]	0.001
Peak ICP (mmHg)	26 [20–32]	34 [24–72]	26 [20–30]	0.008
Number of ICP readings	161 [102–166]	77 [42–124]	163 [122–167]	<0.001
ICP (mmHg)	14 (±12)	26 (±24)	12 (±7)	<0.001
Hospital length of stay (days)	17.0 [11.0–29.0]	5.0 [2.0–8.8]	21.0 [14.0–31.0]	<0.001

[1]*Median [IQR]; Frequency (%); Mean (±SD).*

[2]*Wilcoxon rank-sum test; Pearson chi-squared test; Fisher's exact test; two-sample t-test.*

Table 2. Patient characteristics stratified by in-hospital mortality. PELOD-2 (Pediatric Logistic Organ Dysfunction-2) score is a descriptive scoring system used to evaluate the severity of organ dysfunction. GCS, Glasgow Coma Scale; ICP, intracranial pressure.

Table 3 EMR-based Pediatric Intensity Level of Therapy (ePILOT) scale.	
Variable	**Score**
CSF drainage	1
HyperOsm bolus	1
HyperOsm gtt	1
Neuromuscular blockade	1
$PaCO_2$ 30–35 mmHg*	1
Analgesia/sedation bolus	2
EVD present	2
Pentobarbital gtt	2
$PaCO_2$ < 30 mmHg*	3
Sodium 150–160 mEq/L**	3
Temperature 32–34°C	4
Sodium > 160 mEq/L**	7
Maximum Score	**24**

Excludes OR interventions.
*All therapies applied are summed arithmetically except: * $PaCO_2$ contributes a maximum of 3; ** sodium a maximum of 7.*

Table 3. EMR-based Pediatric Intensity Level of Therapy (ePILOT) scale. CSF, cerebrospinal fluid; EVD, external ventricular drain; HyperOsm, hyperosmolar; gtt, continuous infusion; Na, sodium; OR, operating room; $PaCO_2$, partial pressure of arterial carbon dioxide.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ePILOTSupplement.docx