Development and external validation of a machine learning model for the early prediction of doses of harmful intracranial pressure in patients with severe traumatic brain injury

Development and external validation of a machine learning model for the early prediction of doses of harmful intracranial pressure in patients with severe traumatic brain injury.

Giorgia Carra[1],PhD, Fabian Güiza[1],PhD, Ian Piper[2], PhD, Giuseppe Citerio[3], MD, Andrew Maas[4], PhD, Bart Depreitere[5], PhD, Geert Meyfroidt[1], PhD, on behalf of CENTER-TBI High-Resolution ICU (HR ICU) Sub-Study Participants and Investigators*.

1 Clinical Division and Laboratory of Intensive Care Medicine, Department of Cellular and Molecular Medicine, UZ Leuven and KU Leuven, Herestraat 49, 3000 Leuven, Belgium

2 Intensive Care Monitoring research and Paediatric Critical Care. Edinburgh Royal Hospital for Children and Young People 50 Little France Crescent, Edinburgh bio Quarter Edinburgh, EH16 4TJ

3 University of Milan-Bicocca, School of Medicine and Surgery and Department of Emergency and Intensive Care, Monza, Italy

4 Department of Neurosurgery, Antwerp University Hospital and University of Antwerp, Wilrijkstraat 10, 2650 Edegem, Belgium

5 Department of Neurosurgery, UZ Leuven and KU Leuven, Herestraat 49, 3000 Leuven, Belgium

Corresponding author:	**Prof. Dr. Geert Meyfroidt**
	email: geert.meyfroidt@uzleuven.be
	Phone: +32 16344021
ORCID code corresponding author	0000-0003-4259-3935
Tables: 3	
Figures: 2	
Word count number:	3554

Contact information:

Giorgia Carra	giorgia.carra@kuleuven.be	+32 16328120
Fabian Güiza	fabian.guiza@kuleuven.be	+32 16344021
Ian Piper	ian.piper@brainit.org	+44 1412012595
Giuseppe Citerio	giuseppe.citerio@unimib.it	+39 2334335
Andrew Maas	andrew.maas@uza.be	+32 38213328
Bart Depreitere	bart.depreitere@kuleuven.be	+32 16344290
Geert Meyfroidt	geert.meyfroidt@uzleuven.be	+32 16344021

*CENTER-TBI High Resolution (HR ICU) Sub-Study Participants and Investigators:

Audny Anke [1], Ronny Beer [2], Bo-Michael Bellander[3], Erta Beqiri[4], Andras Buki[5], Manuel Cabeleira[6], Marco Carbonara[7], Arturo Chieregato[4], Giuseppe Citerio[8, 9], Hans Clusmann[10], Endre Czeiter[11], Marek Czosnyka[6], Bart Depreitere[12], Ari Ercole[13], Shirin Frisvold[14], Raimund Helbok[2], Stefan Jankowski[15], Daniel Kondziella[16], Lars-Owe Koskinen[17], Ana Kowark[18], David K. Menon[13], Geert Meyfroidt[19], Kirsten Moeller[20], David Nelson[3], Anna Piippo-Karjalainen[21], Andreea Radoi[22], Arminas Ragauskas[23], Rahul Raj[21], Jonathan Rhodes[24], Saulius Rocka[23], Rolf Rossaint[18], Juan Sahuquillo[22], Oliver Sakowitz[25, 26], Peter Smielewski[6], Nino Stocchetti[27], Nina Sundström[28], Riikka Takala[29], Tomas Tamosuitis[30], Olli Tenovuo[31], Andreas Unterberg[26], Peter Vajkoczy[32], Alessia Vargiolu[8], Rimantas Vilcinis[33], Stefan Wolf[34], Alexander Younsi[26], Frederick A. Zeiler[13,35]

[1] Department of Physical Medicine and Rehabilitation, University hospital Northern Norway
[2] Department of Neurology, Neurological Intensive Care Unit, Medical University of Innsbruck, Innsbruck, Austria
[3] Department of Neurosurgery & Anesthesia & intensive care medicine, Karolinska University Hospital, Stockholm, Sweden
[4] NeuroIntensive Care, Niguarda Hospital, Milan, Italy
[5] Department of Neurosurgery, Medical School, University of Pécs, Hungary and Neurotrauma Research Group, János Szentágothai Research Centre, University of Pécs, Hungary
[6] Brain Physics Lab, Division of Neurosurgery, Dept of Clinical Neurosciences, University of Cambridge, Addenbrooke's Hospital, Cambridge, UK
[7] Neuro ICU, Fondazione IRCCS Cà Granda Ospedale Maggiore Policlinico, Milan, Italy
[8] NeuroIntensive Care Unit, Department of Anesthesia & Intensive Care, ASST di Monza, Monza, Italy
[9] School of Medicine and Surgery, Università Milano Bicocca, Milano, Italy
[10] Department of Neurosurgery, Medical Faculty RWTH Aachen University, Aachen, Germany
[11] Department of Neurosurgery, University of Pecs and MTA-PTE Clinical Neuroscience MR Research Group and Janos Szentagothai Research Centre, University of Pecs, Hungarian Brain Research Program (Grant No. KTIA 13 NAP-A-II/8), Pecs, Hungary
[12] Department of Neurosurgery, University Hospitals Leuven, Leuven, Belgium
[13] Division of Anaesthesia, University of Cambridge, Addenbrooke's Hospital, Cambridge, UK
[14] Department of Anesthesiology and Intensive care, University Hospital Northern Norway, Tromso, Norway
[15] Neurointensive Care, Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, UK
[16] Departments of Neurology, Clinical Neurophysiology and Neuroanesthesiology, Region Hovedstaden Rigshospitalet, Copenhagen, Denmark
[17] Department of Clinical Neuroscience, Neurosurgery, Umeå University, Umeå, Sweden
[18] Department of Anaesthesiology, University Hospital of Aachen, Aachen, Germany
[19] Intensive Care Medicine, University Hospitals Leuven, Leuven, Belgium
[20] Department Neuroanesthesiology, Region Hovedstaden Rigshospitalet, Copenhagen, Denmark
[21] Helsinki University Central Hospital, Helsinki, Finland
[22] Department of Neurosurgery, Vall d'Hebron University Hospital, Barcelona, Spain
[23] Department of Neurosurgery, Kaunas University of technology and Vilnius University, Vilnius, Lithuania
[24] Department of Anaesthesia, Critical Care & Pain Medicine NHS Lothian & University of Edinburg, Edinburgh, UK
[25] Klinik für Neurochirurgie, Klinikum Ludwigsburg, Ludwigsburg, Germany
[26] Department of Neurosurgery, University Hospital Heidelberg, Heidelberg, Germany
[27] Department of Pathophysiology and Transplantation, Milan University, and Neuroscience ICU, Fondazione IRCCS Cà Granda Ospedale Maggiore Policlinico, Milano, Italy
[28] Department of Radiation Sciences, Biomedical Engineering, Umea University, Umea, Sweden

[29] Perioperative Services, Intensive Care Medicine, and Pain Management , Turku University Central Hospital and University of Turku, Turku, Finland
[30] Neuro-intensive Care Unit, Kaunas University of Health Sciences, Kaunas, Lithuania
[31] Rehabilitation and Brain Trauma, Turku University Central Hospital and University of Turku, Turku, Finland
[32] Neurologie, Neurochirurgie und Psychiatrie, Charité – Universitätsmedizin Berlin, Berlin, Germany
[33] Department of Neurosurgery, Kaunas University of Health Sciences, Kaunas, Lithuania
[34] Department of Neurosurgery, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany
[35] Section of Neurosurgery, Department of Surgery, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, MB, Canada

Abstract

Treatment and prevention of elevated intracranial pressure (ICP) is crucial in patients with severe traumatic brain injury (TBI). Elevated ICP is associated with secondary brain injury, and both intensity and duration of an episode of intracranial hypertension, often referred to as “ICP dose”, are associated with worse outcomes. Prediction of such harmful episodes of ICP dose could allow for a more proactive and preventive management of TBI, with potential implications on patients’outcomes. The goal of this study was to develop and validate a machine-learning (ML) model to predict potentially harmful ICP doses in patients with severe TBI. The prediction target was defined based on previous studies and included a broad range of doses of elevated ICP that have been associated with poor long-term neurological outcomes. ML models were used, with minute-by-minute ICP and mean arterial blood pressure signals as inputs. Harmful ICP episodes were predicted with a 30 minutes forewarning. Models were developed in a multi-center dataset of 290 adult patients with severe TBI and externally validated on 264 patients from the Collaborative European Neuro-trauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI) dataset. The external validation of the prediction model on the CENTER-TBI dataset demonstrated good discrimination and calibration (AUC: 0.94, accuracy: 0.89, precision: 0.87, sensitivity: 0.78, specificity: 0.94, calibration-in-the-large: 0.03, calibration slope: 0.93). The proposed prediction model provides accurate and timely predictions of harmful doses of ICP on the development and external validation dataset. A future interventional study is needed to assess whether early intervention on the basis of ICP dose predictions will result in improved outcomes.

Key words for indexing: Traumatic Brain Injury, Intracranial Pressure, Intracranial pressure dose, Machine Learning, Prediction

Introduction

Current guidelines for the management of intracranial hypertension in patients with severe traumatic brain injury (TBI) suggest to initiate treatment when the intracranial pressure (ICP) rises above 22mmHg.[1,2] This threshold-based strategy is population-derived, therefore it does not allow therapy to be targeted to specific subgroups of patients. In addition, secondary injury by elevated ICP may not be adequately defined by the simple crossing of a universal threshold.

The ICP dose, *i.e.* the combination of intensity and duration of an ICP event, might offer a better representation of the risk of secondary brain injury due to elevated ICP. High doses of elevated ICP have been associated with worse clinical outcomes in several observational studies.[3–5] Moreover, the association between ICP doses and long-term neurological outcomes was visualized in a color coded heat map by Güiza et al.[6] (Figure 1 panel C shows an adapted version), which was further replicated in other large datasets.[7,8] In the visualizations, an exponential line separates the ICP doses that occur more frequently in patients with worse and better long-term neurological outcomes, represented in the visualizations in red and blue respectively. These studies corroborated the hypothesis that elevated ICP can be tolerated if maintained for a short period, whereas ICP values between 15mmHg and the current treatment threshold of 22mmHg, if maintained for a prolonged time, resulted associated with poor long-term neurological outcomes. Panel A) and B) of Figure 1 show an example of how the quantification of ICP harmfulness may vary according to the criterion in use, namely whether we rely on the concept of ICP > 22 mmHg or the concept of harmful ICP doses.[6–8]

Despite these scientific evidences, the concept of ICP dose has not been integrated in the clinical reasoning. One potential limitation is the ICP dose retrospective calculation, which as such only provides information on the neurological burden of past events of elevated ICP but gives little information that can be used for prophylactic purpose. A quantification of the risk of the patient to experience impending events of harmful ICP doses could serve this scope and provide valuable information for the attending physician.

In this study, we hypothesized that the analysis of routinely monitored signals through advanced machine learning (ML) techniques, could allow for the early detection of events of harmful ICP doses. ML algorithms use mathematical rules to capture patterns in the observed data and then apply such patterns on a new, unseen dataset. Specifically, the goal of this study was the development and external validation of a ML model to predict a broad range of future harmful ICP doses with a 30 minute forewarning.

Material and methods

Database

The development cohort included the data of 290 patients from 6 prospectively and retrospectively collected databases: the Brain-IT[9] is a European multi-center database that contains data of 206 adult patients with severe TBI admitted to 22 intensive care units (ICU)s between March 2003 and July 2005. Ethical approval for the collection and later analysis of the data was obtained from the Multi-Centre Research Ethics Committee for Scotland MREC/02/09. Ethical approval was additionally obtained from the local medical ethics committee of the centers involved. Of the remaining 84 patients: 38 patients were admitted in the San Gerardo Hospital in Monza, Italy between March 2010 and April 2013; 27 patients from the University Hospitals of Leuven, Belgium between September 2010 and September 2013; 19 patients from the NEMO (Individualized targeted monitoring in neurocritical care) project at the Antwerp University Hospital, Belgium, between March 2010 and June 2013; All centers obtained ethical approval from the local medical ethics committee.

The external validation cohort was composed by 264 patients included in the High Resolution substudy of the Collaborative European Neuro-trauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI) dataset[10]. The CENTER-TBI dataset prospectively collects data of adult patients with TBI admitted to 47 European ICUs between 2015 and 2017. Ethical approval for CENTER-TBI was obtained from the local ethic committee for each recruiting site.

All datasets include continuous (minute-by-minute) recordings of ICP and MAP signals. Missing data of duration less than two consecutive values were imputed with the median value of the previous ten minutes recordings. Patients were declared eligible for the study if their ICP recordings were acquired with an intra-parenchymal ICP probe. Intracranial hypertension was treated according to the guidelines for the treatment of severe TBI[11] in force during data acquisition.

Predictive task

The model provides the probability that the patient will experience, in the next 30 minutes, an event of ICP dose that appeared to be associated with poor long-term neurological outcomes in the visualization method proposed by Güiza et al.[6], and in the following validation studies.[7,8] Specifically, the prediction target is displayed with the yellow line in Figure 1 Panel D.

Model development

The prediction model for a broad range of potentially harmful ICP doses was developed in 2 subsequent steps.

Given that the predictive patterns that precede harmful ICP doses may differ across the red area, to obtain optimal performance we divided the red area into several sub-areas and targeted each sub-area separately. The red sub-areas were defined according to the visualization curves[6–8] as follows: ICP > 15 mmHg for more than 180 minutes, ICP > 18 mmHg for more than 70 minutes, ICP > 20 mmHg for more than 35 minutes, ICP > 22 mmHg for more than 25 minutes, ICP > 24mmHg for more than 18 minutes, ICP > 26 mmHg for more than 14 minutes, ICP > 28 mmHg, ICP > 30 mmHg, and ICP > 34 mmHg for more than 10 minutes. An example of the red subareas can be seen in Figure 1, panel D. In the first step, we developed a specific prediction model for ICP doses belonging to each identified sub-area. Each model for the prediction of the red sub-areas was based on a Gaussian Regressor model with a Rational Quadratic kernel function. We will refer to these sub-

models as GP_X, where X is the lower ICP threshold that identifies that specific sub-area. For example, GP_{15} refers to the prediction model for the red sub-area delineated by ICP doses of ICP > 15 mmHg for more than 180 minutes, see Figure 1 panel D. Importantly, for the GP_{30} model we used the prediction model previously proposed by Güiza et al. .[12] Input features for the GPx models were extracted from the 4 hours of continuous ICP, MAP and LAx signals preceding the prediction, namely between t(-239) and t(0) as shown in the example of Figure 1 panel B. For a complete list of the extracted features see the Supplementary Material. To avoid overfitting, for each GP model, the most predictive features were selected through the combination of a linear and non-linear method, i.e. feature selection via LASSO[13] and features selection via mutual information.[14] Feature importance was computed with the permutation importance technique.[15] Hyper-parameters tuning was performed for each GPx model separately.

In a second phase, to minimize the prediction error of the single sub-models and provide to the clinicians a unique model output, we developed a Random Forest (RF) classifier that combines the predictions of the models for the red sub-areas and provides as single output the probability that the patient will experience events of ICP in the red area in the next 30 minutes. For simplicity, we will further refer to this model as RFred model. We believe this output type, *i.e.* probability of being in the red area, is particularly congenial to the clinical environment. For this task, the RF classifier demonstrated superior performance as compared to more simple models (linear regression model, GP classifier, and decision tree classifier). The only inputs to the RFred model are the predictions of the GPx models for the red sub-areas. No ICP, MAP or LAx features were entered into the RFred model to avoid information leakage.

The use of a model architecture at 2 layers (GPx and RF) demonstrated better performance as compared to the alternative approaches of using one model to predict directly the entire red area or not using a ML model to combine the predictions of the single GPx models. More information on the model construction can be found in the Supplementary Material.

To increase generalizability and avoid overfitting, models were trained with 10-fold cross validation (CV).

External validation

The GPx_s and RFred models were externally validated on the 264 patients from the CENTER-TBI dataset.[10] External validation is a crucial step in model development, it assesses the model's generalizability capacities and provides an estimate of the future performance of the model when applied to an unknown population of patients. External validation is even more important in this study to assess whether changes in clinical practice may affect the performance of the model, given that the development set is more than 10 years-old.

Model performance and statistical analysis

Performance of the models was assessed with the following metrics: area under the receiver operating characteristic curve (AUC), area under the precision-recall curve (AP), accuracy, precision, sensitivity, and specificity. Performance metrics on the development cohort were provided in terms of mean (SD) across the 10-fold CV iterations. The calibration was assessed by using calibration plots and by computing the calibration-in-the-large and calibration-slope. Clinical importance was assessed with decision curves. Decision curve analysis compares the clinical usefulness of using the prediction model to alert the clinicians (and therefore trigger medical interventions) with the opposite strategies of "alert for all" or "alert for none". The "alert for all" indicates the scenario in which the clinician would be constantly evaluating the clinical situation of the patient, while the "alert for none" indicates the implausible scenario in which the clinician would never be alerted by the condition of the patient. For this prediction model, medical intervention represents the need for an additional medical evaluation of the clinical status of the patient.

Analyses were performed in Python (*version 3.5,* https://www.python.org/) with the following libraries: numpy (*version 1.15,* https://numpy.org/), sklearn (*version 1.1*, https://scikit-learn.org/stable/) and scipy (*version 0.20*, https://www.scipy.org/). Calibration curves were extracted with the R-based library givitiR (*version 1.3*, https://CRAN.R-project.org/package=givitiR).

Trustiness and transparency of the model

This study adheres to the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) reporting guideline.[16] To favor trustiness and transparency, we provided a model fact sheet that summarizes the main characteristics of the proposed model, after the example proposed by Brajer et al. .[17]

Results

Patients' demographics characteristics are reported in Table 1. Patients experienced a median [IQR] number of events of ICP dose in the red area of 5 [1-28], for a median [IQR] percentage of monitoring time spent in the red area of 12% [0-46], against a median [IQR] percentage of monitoring time spent with ICP > 22 mmHg of 1% [0-6]. The external validation cohort included 8421 events of ICP dose in the red area and 16840 selected events in the blue area.

Here we only report the results of the external validation. In short, for what concerns the performance on the development cohort, on the 10 folds CV internal validation sub-sets the RFred presented a mean (SD) AUC of 0.92 (0.02), AP of 0.87 (0.03), accuracy of 0.86 (0.02), precision of 0.81 (0.04), sensitivity of 0.76 (0.04) and specificity of 0.91 (0.02). More detailed performance of the models on the development cohort and feature importance analysis can be found in the Supplementary Material.

On the CENTER-TBI dataset, all GPx models presented an AUC above 0.83, an AP above 0.75, an accuracy above 0.74, a precision above 0.57, a sensitivity above 0.55, and specificity above 0.71. The GP models presented a calibration-in-the-large below 0.05 and mean calibration slope between 0.78 and 1.10. The calibration curves p-values were below 0.049. See Table 2 for complete results for each model.

When tested on the CENTER-TBI dataset, the RFred model for the prediction of the complete red area presented an AUC of 0.94, an AP of 0.90, an accuracy of 0.89, a precision of 0.87, a sensitivity of 0.78, and a specificity of 0.94. Visually, the model showed adequate calibration, with a calibration-in-the-large of 0.03 and a calibration slope of 0.91 (despite a p-value < 0.01), see Figure 2, panel A. Also on

the CENTER-TBI dataset the model presented higher clinical benefit than the “alert for all” and “alert for none” options in the risk range [0.12 to 0.87], see Figure 2, panel B. Table 3 summarizes the main characteristics of the model in the form of a model fact sheet.

Discussion

In this study, we present a ML model for the early detection of potentially harmful doses of ICP in patients with severe TBI. The model, which predicts a broad range of ICP doses previously associated with poor long-term neurological outcomes, has good performance and good clinical utility even when validated on an external, multi-center, prospectively collected dataset.

In the past, several studies have attempted to predict single ICP values or episodes of elevated ICP,[12,18–20] with common characteristics. Some of these models focused on one specific ICP insults of specific intensity and duration, not taking into consideration the complex, broad range of ICP events that have been associated with poor outcomes.[12,18] Some of these prediction models use a short forewarning time, which may be insufficient to trigger a clinically useful intervention.[19,20] An additional characteristic that could challenge clinical implementation is the large number of required inputs, often from multiple monitoring sources, which not only obstacles clinical implementation but also increases the risk of overfitting.[19] Last, but most importantly, most of these models lack external validation on geographically and temporally independent datasets.[18–20] External validation is strongly recommended,[16] to assess the model generalizability capacities and consequently to evaluate the performance of the model when applied to a general, unknown population.

The present model presents an answer to these issues in many ways.

First, the model presents good performance also when externally validated on the CENTER-TBI dataset, with an AUC of 0.94, an AP of 0.90, an accuracy of 0.89, a precision of 0.87, a sensitivity of 0.78, and a specificity of 0.94. As this large, external, multicenter dataset was collected more than ten years after the development cohort, this good performance not only proves the robustness of the model towards its

application to different ICU settings, but it also suggests robustness to changes in the clinical practice. On the CENTER-TBI dataset the model presented clinical usefulness within the risk thresholds [0.12-0.87]. Acceptable alerting thresholds will need to be evaluated carefully by the clinician and will depend on the risk level of the medical intervention that may be triggered by the alert. In other words, the alerting threshold will depend on how much the clinician accepts an high number of false positives (high sensitivity) as compared to a high number of false negatives (high specificity).

Second, the prediction target is a broad range of episodes of doses of intracranial hypertension.[6] This target was based on a previous study,[6] but similar associations with outcome were observed in a large single-center cohort[7] and in the CENTER-TBI dataset.[8] This broad target provides a more complete approach to the prevention of potentially harmful doses of ICP, targeting ICP events that are not only associated with increased mortality but that are also associated with reduced neurological outcomes. Moreover, this model represent a practical step towards the use of the concept of "ICP dose" for prophylactic purpose at the bedside.

Third, the forewarning time interval was defined after consultation with three clinicians of the ICU of the University Hospitals of Leuven, Belgium, and a 30 minutes forewarning was identified as adequate to trigger a useful clinical response.

Fourth, and finally, the model is sparse, given that it requires as inputs only the continuous ICP and MAP signals, two signals that are routinely recorded in patients with severe TBI.

This study has some limitations. First, harmful ICP events were based on the visualization proposed by Guiza et al.[6] and on the visualizations obtained in further validation studies.[7,8] Although the visualizations obtained in the different datasets have different transition curves, the prediction target of the presented model was associated with worse long-term outcomes in all the studies. Second, the color-coded visualization curves describe the association between ICP doses and long term neurological outcomes on the general population. Nevertheless, the color-coded visualization may vary for different sub-groups of patients (males vs females, old vs young, low vs high treatment intensity level etc..).[6] To date, obtaining such visualizations for stratified groups of patients with TBI is challenging, since the

necessary large, preferably prospective, datasets with large subgroups of patients satisfying the condition of interest are lacking. Until such datasets become available, the development of predictions models for different sub-groups of patients remains challenging. Third, the data used were all acquired in European settings. Validation on datasets collected in non-European ICUs are therefore necessary to fully assess generalizability. Fourth, the CAR status of the patient can affect the association between ICP dose and outcomes, where patients can better tolerate elevated doses of ICP if CAR is preserved.[21] In an ideal setting, the prediction target of the model should dynamically adapt to changes in the CAR status. However, optimal methods to measure CAR have not been developed yet. Therefore, in this study we limited the prediction target to the case unadjusted for CAR status. Information on the CAR status of the patient was included as input to the models in the form of LAx signal. The fifth limitation is linked to the lack of absolute explainability of ML models, which may limit the degree to which it is trusted and accepted by clinicians. To overcome this limitation, we performed a feature importance analysis for each GP_x model. In addition, to increase the trustiness and transparency of the model, not only did we report the results in accordance with the TRIPOD guidelines, but we also provided a model fact sheet to summarize the main characteristics of the proposed model. Sixth, the models uses ICP and MAP features that are extracted from the previous 4 hours of ICP monitoring, which implies that in a clinical setting, clinicians could receive the first prediction only 4 hours after the start of ICP monitoring. Although this may represent a source of delay, given the average length of monitoring of these patients and that the minimum recommended duration of ICP monitoring is of 72 hours,[2] we believe that the model could still provide useful information to the clinicians. Seventh, the relatively small size of the dataset may lead to a slight over-representation of certain patients in the development of the GPx models. However, the good performance that were obtained in the external validation dataset shows that the risk of overfitting was minimal or even neglectable. Eight, the model was developed on data of treated patients although information on treatment strategy was not included in the model. Patient management may play a confounding role in the model development and evaluation. This is an intrinsic limitation of the study, which could be partially overcome by adding high-resolution treatment information in the model. Such high-resolution information was not available in our development cohort, moreover it remains unclear whether the inclusion of such information may limit the translation of such model at

the bedside (where summarized information about patient management would need to be automatically computed and input in the model). The last limitation of the study is that we target ICP doses that have been associated with poor outcomes. However, association does not infer causality, and whether early intervention on the basis of predicted doses of ICP will result in improved outcomes needs to be assessed in future interventional studies.

Despite these limitations, the presented model represents the first example of an accurate model for the prediction of a broad range of harmful ICP doses in patients with TBI. To develop and externally validate this model we used two of the largest multicenter databases for patients with TBI with continuous monitoring data and outcomes. Further model validation on more recent multicenter data may be required. We believe that this model could provide valuable information for the clinical management of patients with TBI. Future studies will focus on evaluating the performance of the model when prospectively applied on continuous signals at the bedside. In context, of particular interest is the impact of treatment intensity for elevated ICP on predictions, which needs to be further evaluated. In addition, future randomized clinical trials will be necessary to identify the risk of medical interventions that may be taken in response to the alerts of the models, to assess user acceptance and most importantly to assess whether the use of this model at the bedside may have a positive impact on time spent by the patient with an ICP dose in the red area.

Conclusions

In this study, we present an accurate and robust model for the early detection of events of ICP doses that are associated with worse long-term neurological outcomes. Using only the ICP and MAP signals, our model can predict with 30 minutes forewarning, events of harmful doses of ICP with high accuracy, high sensitivity and specificity. The model presents good performance even when validated on a large external multicenter dataset, showing robustness to changes in clinical setting and practice. Future interventional studies are needed to assess the impact of the use of this tool at the bedside on clinical practice and on patient outcomes.

Funding information

Research Foundation Flanders (FWO), internal KU Leuven funds (C2 project [C24/17/072]). The collection of the BrainIT dataset was funded by the FP5 and FP7 European Union programs (grants: FP5 QLRI-2000-00454, FP5 QLGT-2002-00160, FP7 IST-2007-217049). The NEMO project was funded by the Flemish Government Agency for Innovation by Science and Technology (IWT)—Applied Biomedical Research (TBM) Program (project 080660).The collection of the CENTER-TBI Database was supported by the FP7 program of the European Union (grant 602150). Additional funding for CENTER was obtained from the Hannelore Kohl Stiftung (Germany), from OneMind (USA), from Integra LifeSciences Corporation (USA), and from NeuroTrauma Sciences (US). Carra performed this research while receiving funding from the Research Foundation, Flanders (FWO) as a PhD fellow.

Author disclosure statement

Maas receives funding from Fp7 program EU, Hannelore Kohl Stiftung and Integra Life Sciences. He is part of a Data Safety Monitoring Board for PressuraNeuro.

All the other co-authors declare no conflicts of interest.

Author contribution statement

GM*, FG*, GCa*, BD designed and conceptualized the study. Data acquisition was supervised by GCi, AM, IP and by the CENTER-TBI high-resolution (HR-ICU) sub-study participants and investigators. Design of the methodology for data analysis, formal data analysis and interpretation of the results was performed by GCa, FG and GM. GCa drafted the manuscript, which was later revised by all authors. GM and FG supervised the study, GM was the principal investigator of the study.

* Contributed equally

References

1. Carney N, Totten AM, O'Reilly C, et al. Guidelines for the Management of Severe Traumatic Brain Injury 4th Edition. 2016;80. https://doi.org/10.1227/neu.0000000000001432.

2. 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;45:1783–1794. https://doi.org/10.1007/s00134-019-05805-9.

3. Flechet M, Güiza F, Vlasselaers D, et al. Near-infrared cerebral oximetry to predict outcome after pediatric cardiac surgery: A prospective observational study. Pediatr Crit Care Med. 2018;19:433–441. https://doi.org/10.1097/pcc.0000000000001495.

4. Lazaridis C, DeSantis SM, Smielewski P, et al. Patient-specific thresholds of intracranial pressure in severe traumatic brain injury. J Neurosurg. 2014;120:893–900. https://doi.org/10.3171/2014.1.jns131292.

5. Kahraman S, Dutton RP, Hu P, et al. Automated Measurement of "Pressure Times Time Dose" of Intracranial Hypertension Best Predicts Outcome After Severe Traumatic Brain Injury. J Trauma Inj Infect Crit Care. 2010;69:110–118. https://doi.org/10.1097/ta.0b013e3181c99853.

6. Güiza F, Depreitere B, Piper I, et al. Visualizing the pressure and time burden of intracranial hypertension in adult and paediatric traumatic brain injury. Intensive Care Med. 2015;41:1067–1076. https://doi.org/10.1007/s00134-015-3806-1.

7. Donnelly J, Güiza F, Depreitere B, Meyfroidt G, Czosnyka M, Smielewski P. Visualising the pressure-time burden of elevated intracranial pressure after severe traumatic brain injury: a retrospective confirmatory study. Br J Anaesth. 2021;126:e15–e17. https://doi.org/10.1016/j.bja.2020.09.018.

8. Åkerlund CA, Donnelly J, Zeiler FA, et al. Impact of duration and magnitude of raised intracranial pressure on outcome after severe traumatic brain injury: A CENTER-TBI high-resolution group study. Kobeissy FH, editor. PLoS One. 2020;15:e0243427. https://doi.org/10.1371/journal.pone.0243427.

9. Piper I, Citerio G, Chambers I, et al. The BrainIT group: concept and core dataset definition. Acta Neurochir (Wien). 2003;145:615–629. https://doi.org/10.1007/s00701-003-0066-6.

10. Maas AIR, Menon DK, Steyerberg EW, et al. Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI). Neurosurgery. 2015;76:67–80. https://doi.org/10.1227/neu.0000000000000575.

11. Carney N, Totten AM, O'Reilly C, et al. Guidelines for the Management of Severe Traumatic Brain Injury, Fourth Edition. Neurosurgery. 2016;80:6–15. https://doi.org/10.1227/neu.0000000000001432.

12. Güiza F, Depreitere B, Piper I, et al. Early Detection of Increased Intracranial Pressure Episodes in Traumatic Brain Injury: External Validation in an Adult and in a Pediatric Cohort. Crit Care Med. 2017;45:e316–e320. https://doi.org/10.1097/ccm.0000000000002080.

13. Tibshirani R. Regression Shrinkage and Selection via the Lasso. J R Stat Soc. 1996;58:267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

14. Frénay B, Doquire G, Verleysen M. Is mutual information adequate for feature selection in regression? Neural Networks. 2013;48:1–7. https://doi.org/10.1016/j.neunet.2013.07.003.

15. Leo Breiman. Random Forests. Mach Learn. 2001;45:5–32.

https://doi.org/10.1023/a:1010933404324.

16. Moons KGM, Altman DG, Reitsma JB, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): Explanation and elaboration. Ann Intern Med. 2015;162:W1–W73. https://doi.org/10.7326/m14-0698.

17. Brajer N, Cozzi B, Gao M, et al. Prospective and External Evaluation of a Machine Learning Model to Predict In-Hospital Mortality of Adults at Time of Admission. JAMA Netw Open. 2020;3:e1920733. https://doi.org/10.1001/jamanetworkopen.2019.20733.

18. Myers RB, Lazaridis C, Jermaine CM, Robertson CS, Rusin CG. Predicting Intracranial Pressure and Brain Tissue Oxygen Crises in Patients with Severe Traumatic Brain Injury. Crit Care Med. 2016;44:1754–1761. https://doi.org/10.1097/ccm.0000000000001838.

19. Bonds BW, Yang S, Hu PF, et al. Predicting secondary insults after severe traumatic brain injury. J Trauma Acute Care Surg. 2015;79:85–90. https://doi.org/10.1097/ta.0000000000000698.

20. Swiercz M, Mariak Z, Krejza J, Lewko J, Szydlik P. Intracranial Pressure Processing with Artificial Neural Networks: Prediction of ICP Trends. Acta Neurochir (Wien). 2000;142:401–406. https://doi.org/10.1007/s007010050449.

21. Maas AIR, Menon DK, Adelson PD, et al. Traumatic brain injury: integrated approaches to improve prevention, clinical care, and research. Lancet Neurol. 2017;4422. https://doi.org/10.1016/s1474-4422(17)30371-x.

Figures

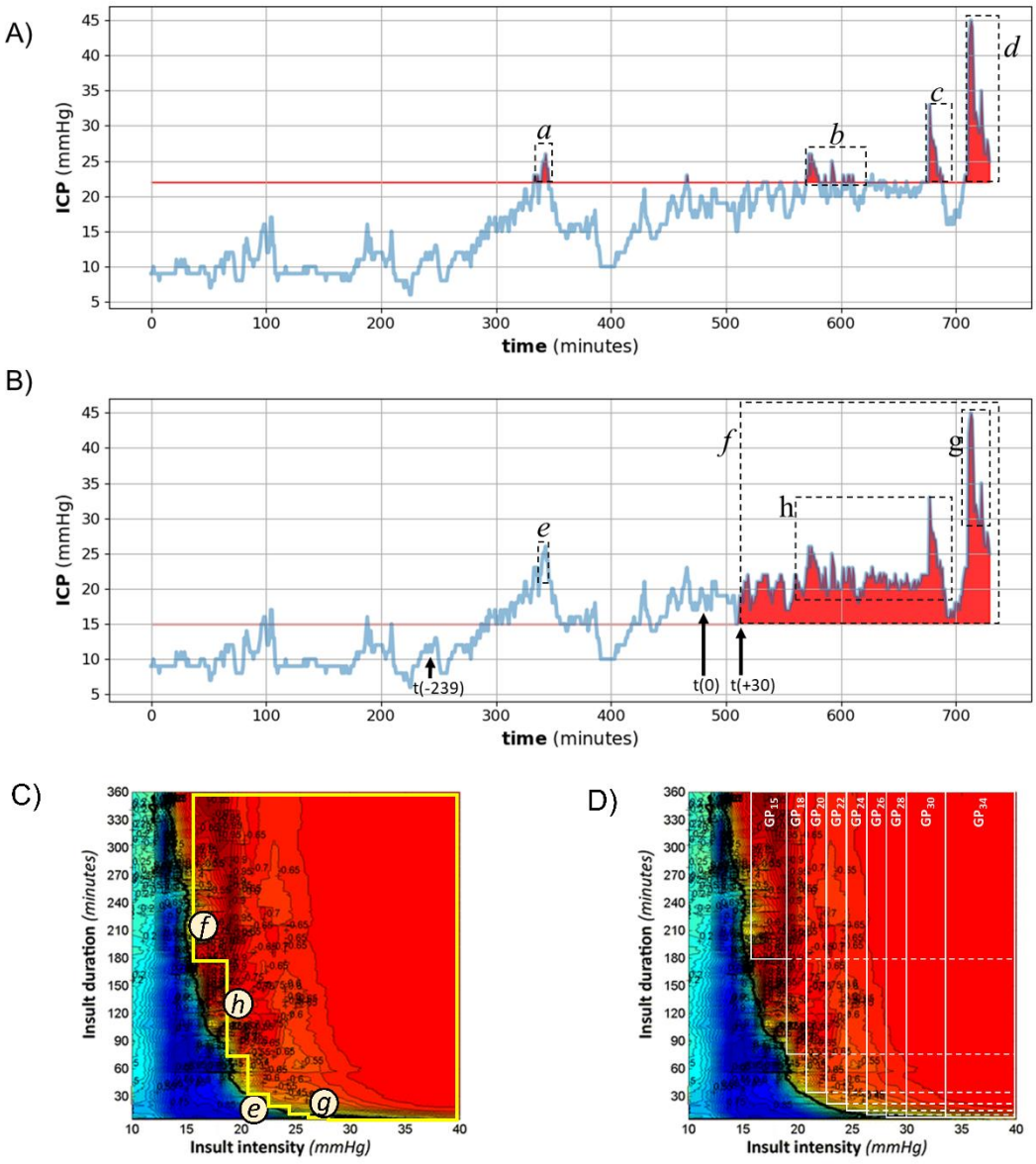

Fig. 1 Panel A) and B) Visualization of intracranial pressure (ICP) harmfulness as evaluated according to different criteria. Panel A) ICP harmfulness is defined by events of ICP > 22mmHg. Events of ICP

that meet this criterion are identified with the letters *a*, *b*, *c* and *d*. Panel B) ICP harmfulness is defined by events of ICP in the "red area" of the visualization proposed by Güiza et al [6] (see panel C). Examples of the events of ICP that meet this criteria are identified with the letters *f*, *g* and *h*. In detail, *f* is an event of ICP>15mmHg that lasts more than 180 minutes, *g* is an event of ICP > 28mmHg that lasts more than 10 minutes, and *h* is an event of ICP > 18mmHg that lasts more than 70 minutes. The same events are also displayed in the visualization of panel C) with white circles. In event *e* the ICP is above 22mmHg but not long enough (less than 25 minutes) to be considered in the "red area" of the visualization. The model provides a prediction at t(0), for impending harmful events at t(+30). The model prediction is based on features that are extracted in the past 4 hours of monitoring, starting form t(-239) to t(0). Panel C) and D) Visualization of the association between doses of ICP, identified by intensity and duration, and the 6-months Glasgow Outcomes Score (GOS). Panel C) shows an adaptation of the original representation as proposed by Güiza et al. [6]. Events of dose of ICP that occur more frequently in patients with worse GOS are represented in red, while events of ICP that occur more frequently in patients with better 6-months GOS are represented in blue. The four white circles indicate the position of the ICP events displayed in panel B). The events of ICP doses that define the prediction target are indicated with the yellow border. Importantly, the prediction target is also part of the "red area" of the color-coded visualizations that were obtained in the following validation studies [7,8]. Panel D) visual representation of the division of the prediction target in sub-areas. Every sub-area extends until the 40mmHg threshold (extreme right border of the figure), where multiple subareas can overlap. Dashed lines indicate when the border of a sub-area overlaps with other sub-areas. A dedicated prediction model was developed for each sub-area. The name of the model for each sub-area is indicated in the top as GPx, where x is the corresponding lower ICP threshold that identifies that specific sub-area.

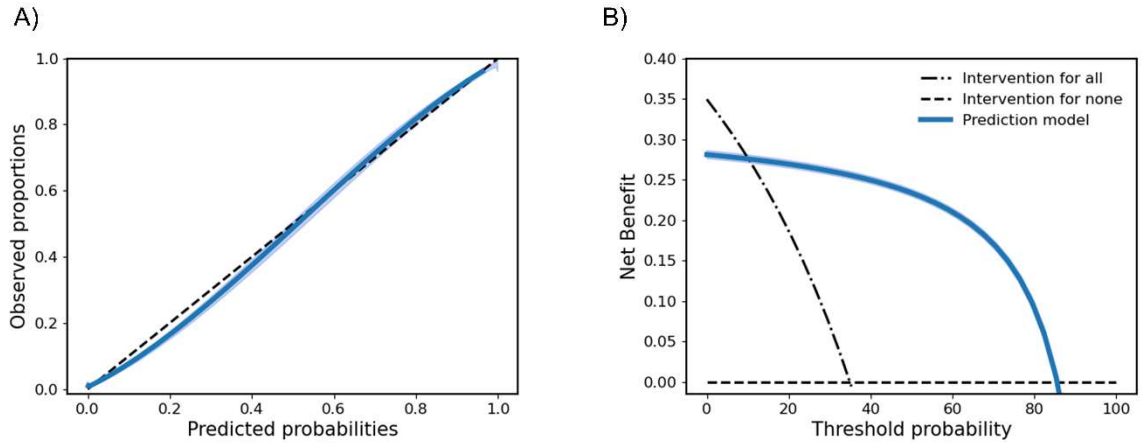

Fig. 2 Performance metrics of the RFred model for the prediction of the red area on the external validation dataset. Panel A) calibration curve. Panel B) decision curve.

Tables

Table 1 Demographics characteristics of the development and validation cohort

Variable	Development cohort	External validation cohort	p-value
Age, years	42 (27 to 56)	47 (29 to 61)	0.04
Sex, male, %	80	81	0.50
GCS at admission	7 (4 to 11)	6 (3 to 10)	0.03
DC, yes, %	16	24	<0.01
Length of monitoring, days	8 (5 – 14)	5 (3 to 6)	<0.01
Number of ICP episodes in the red area	7 (0 to 30)	5 (1 to 28)	<0.01

GCS: Glasgow Coma Scale

DC: Decompressive craniectomy

Table 2 Performance of the GPx and RFred models on the CENTER-TBI dataset.

Model	AUC	AP	Accuracy	Precision	Sensitivity	Specificity	Calibration-in-the-large	Calibration slope
RFred	0.94	0.90	0.89	0.87	0.78	0.94	0.03	0.91
GP_{34}	0.89	0.80	0.86	0.79	0.78	0.90	0.00	1.10
GP_{30}[§]	0.93	0.75	0.88	0.73	0.83	0.91	-0.04	1.22
GP_{28}	0.87	0.80	0.80	0.66	0.80	0.80	0.01	0.96
GP_{26}	0.88	0.81	0.80	0.67	0.78	0.81	0.03	1.03
GP_{24}	0.86	0.80	0.78	0.65	0.78	0.79	0.02	1.06
GP_{22}	0.86	0.78	0.80	0.67	0.76	0.81	0.05	0.98
GP_{20}	0.83	0.75	0.74	0.58	0.80	0.71	0.01	0.89
GP_{18}	0.84	0.76	0.74	0.57	0.84	0.68	0.04	0.78
GP_{15}	0.83	0.77	0.83	0.89	0.55	0.97	0.01	0.84

§ Area under the receiver operating characteristic curve (AUC), accuracy, sensitivity, specificity, calibration-in-the-large and calibration slope were reported from "Development and external validation of a novel algorithm for the early prediction of doses of harmful intracranial pressure in patients with traumatic brain injury", Carra et al., Intensive Care Medicine, 2020. The area under the precision-recall curve (AP) and precision were not reported in the original study from Carra et al. but they were computed specifically for this study.

AUC: Area under the receiver operating characteristic curve; AP: Area under the precision-recall curve.

Table 3 Model fact sheet

Model fact sheet
Model name: Prediction model for harmful ICP doses in patients with traumatic brain injury (TBI).
Summary This model uses intracranial pressure (ICP) and mean arterial blood pressure (MAP) inputs to predict, with a 30 minutes forewarning, a broad range of events of ICP doses associated with worse long-term neurological outcomes in patients with TBI. The model was developed by the Laboratory of Intensive Care Medicine of KU Leuven, Belgium, between 2020 and 2021.
Mechanisms • **Outcome** ……...……. predictions of potentially harmful ICP doses • **Output** …… 0% to 100% probability of a future event of harmful ICP doses • **Suggested alerting thresholds** ……. between 20% and 80%, depending on the risk level of the medical intervention • **Patient population** ……. Adults (>18 years old) with severe TBI and invasive intra-parenchymal ICP monitoring • **Time of prediction** …… 30 minutes before the event onset • **Predictions data type** ……... minute-by-minute predictions • **Input data type** ……. continuous ICP and MAP recordings • **Input data source** ..... bedside monitors or local Patient Data Management System (PDMS) • **Training data size** …… 290 adult patients with severe TBI (20938 samples) • **Model type** …….. ensembled GP-based models and RF-based model
Validation and Performance • **External validation on CENTER-TBI (264 patients):** AUC: 0.94, AP: 0.89, accuracy: 0.88, precision: 0.85, sensitivity: 0.77, specificity: 0.93.
Uses and directions This model is intended to be used as an additional source of information on which to base the management of patients with severe TBI. In specific, this model is intended to be used for the early identification of future events of ICP doses that have been associated with poor long-term neurological outcomes in previous studies.

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation	Page number
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1,4
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	4
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	5
Objectives	3	State specific objectives, including any prespecified hypotheses	5,6
Methods			
Study design	4	Present key elements of study design early in the paper	6,7
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	6
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	6,7
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	n.a.
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	7
Data sources/ measurement	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	*7*

Bias	9	Describe any efforts to address potential sources of bias	n.a.
Study size	10	Explain how the study size was arrived at	n.a.
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	n.a.
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	7 - 9
		(*b*) Describe any methods used to examine subgroups and interactions	n.a.
		(*c*) Explain how missing data were addressed	7
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy	n.a.
		(*e*) Describe any sensitivity analyses	n.a.
Results			
Participants	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	10
		(b) Give reasons for non-participation at each stage	n.a.
		(c) Consider use of a flow diagram	n.a.
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	n.a.
		(b) Indicate number of participants with missing data for each variable of interest	n.a.
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	n.a.
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	*10*
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	n.a.
		Cross-sectional study—Report numbers of outcome events or summary measures	n.a.

Main results	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	n.a.
		(*b*) Report category boundaries when continuous variables were categorized	n.a.
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	n.a.
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	n.a.
Discussion			
Key results	18	Summarise key results with reference to study objectives	11
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	12,13
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	13,14
Generalisability	21	Discuss the generalisability (external validity) of the study results	13,14
Other information			
Funding	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	17,18

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org