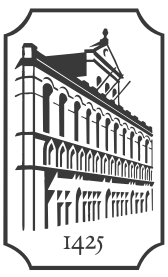
ACTA BIOMEDICA LOVANIENSIA 801
KU Leuven
Biomedical Sciences Group
Faculty of Medicine
Department of Cellular and Molecular Medicine
Laboratory of Intensive Care Medicine

Giorgia Carra

Artificial Intelligence for Decision Support and Treatment Optimization in Patients with Brain Injuries



Leuven University Press

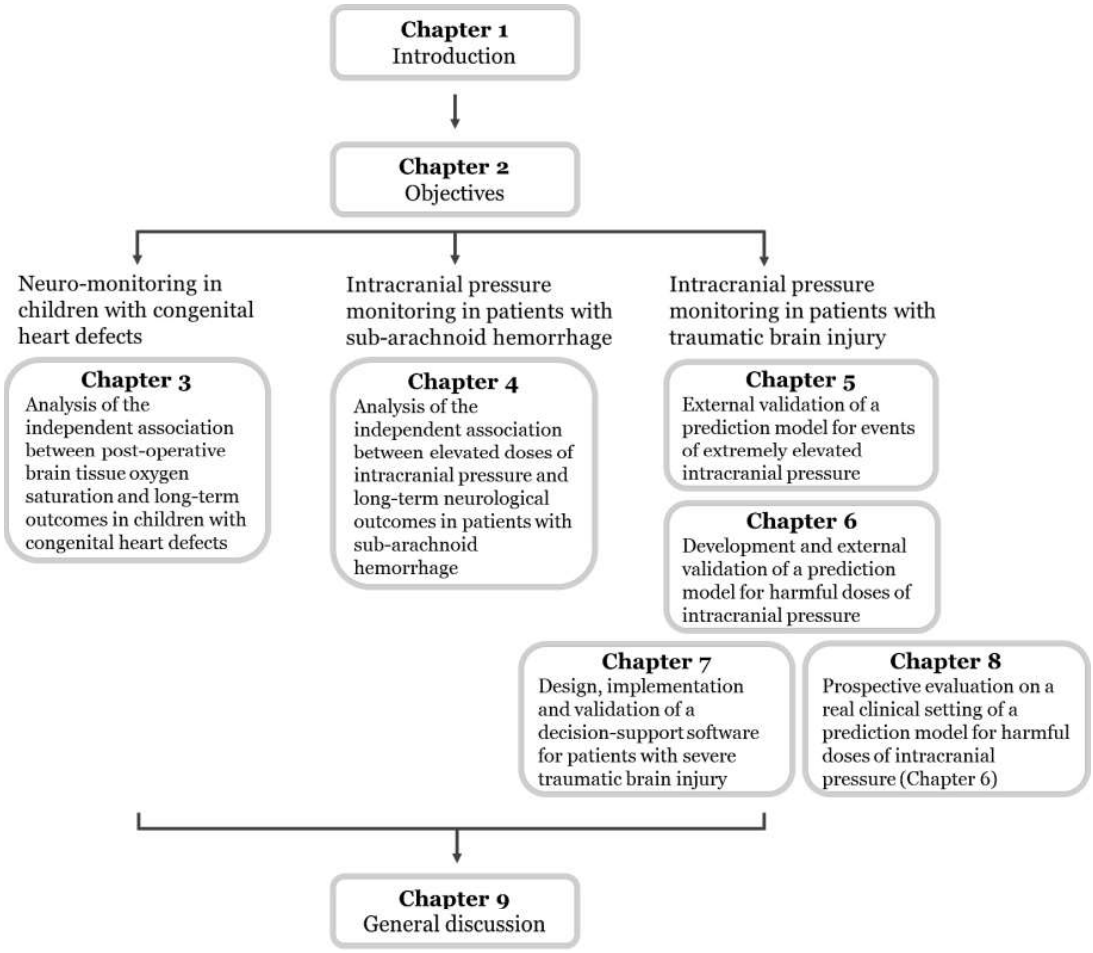
Thesis submitted in partial fulfillment of the requirements for the degree of «Doctor of Biomedical Sciences».

Jury:

Supervisor:	Prof. Dr. Geert Meyfroidt
Co-Supervisors:	Dr. Fabian Güiza
	Prof. Dr. Bart Depreitere
Chair examining committee:	Prof. Dr. Koen Milisen
Chair public defense:	Prof. Dr. Patrizia Agostinis
Jury members	Prof. Dr. Tom Theys
	Prof. Dr. Celine Vens
	Prof. Dr. David Menon
	Prof. Dr. Fabio Taccone

ISBN 978 94 6165 338 3
D/2022/1869/19

CONTENTS



LIST OF ABBREVIATIONS

A	
A	Vargha and Delaney's A
AI	Artificial intelligence
AHA	American heart association
AP	Area under the precision recall curve
AUC	Area under the receiver operating characteristic curve
AUROC	Area under the receiver operating characteristic curve
aSAH	Aneurysmal subarachnoid hemorrage
B	
BMI	Body mass index
C	
CAR	Cerebral autoregulation
CENTER-TBI	Collaborative European neuro-trauma effectiveness research in traumatic brain injury
CHD	Congenital heart defect
CI	Confidence interval
C_RI	Credible interval
CPB	Cardio pulmonary bypass
CPP	Cerebral perfusion pressure
CSF	Cerebrospinal fluid
CT	Computer tomography
CV	Cross-validation
D	
DCI	Delayed cerebral ischemia
DHCA	Deep hypothermic circulatory arrest
DM	Data mining
DT	Decision tree
E	

EHR	Electronic health record
EVD	External ventricular drain
F	
FN	False negatives
FP	False positives
G	
GCS	Glasgow come scale
GDPR	General Data Protection Regulation
GOS	Glasgow outcome score
GP	Gaussian processes
H	
H&H	Hunt and Hess
I	
ICU	Intensive care unit
IQ	Intelligence quotient
IQR	Interquartile range
IT	Internet technology
L	
LAx	Low-frequency autoregulation index
LOM	Length of monitoring
LOS	Length of stay
M	
MAP	Mean arterial blood pressure
MABP	Mean arterial blood pressure
mFisher	Modified Fisher score
ML	Machine learning
K	
KS	Kolmogorov-Smirnov
N	
NB	Net benefit

NCS	Neurocritical care society
NIRS	Near-infrared spectroscopy
NPV	Negative predictive value
O	
OR	Odd ratio
P	
$PbtO_2$	Partial pressure of brain tissue oxygen
PICU	Pediatric intensive care unit
PIM	Pediatric index of mortality
PDMS	Patient data management system
PN	Parenteral nutrition
PPV	Positive predictive value
PRx	Pressure reactivity index
R	
RACHS	Risk adjustment for congenital heart surgery
RCT	Randomized clinical trial
RF	Random forests
RFred	Random forest model for the prediction of harmful ICP doses (Chapter 6)
ROC	Receiver operating characteristic curve
S	
SBI	Secondary brain injury
$SctO_2$	Cerebral tissue oxygen saturation
SD	Standard deviation
SIBICC	Seattle International Severe Traumatic Brain Injury Consensus Conference
SOFA	Sepsis-related organ failure assessment
T	
TCD	Transcranial doppler
TN	True negatives
TP	True positives

TRIPOD	Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (reporting guidelines)
TRL	Technology readiness level
U	
UI	User interface
UZ Leuven	University Hospitals Leuven
W	
WISC	Wechsler intelligence scale for children
WFNS	World Federation of Neurosurgical Societies
WPPSI	Wechsler preschool and primary scale of intelligence

1

INTRODUCTION

PARTIALLY ADAPTED FROM: **Carra G.** et al. Data-driven ICU management: Using big data and algorithms to improve outcomes. Journal of Critical Care. 2020 Dec. 60;300-304.

1.1 CRITICAL ILLNESS

1.1.1 *Overview*

Patients who suffer from life-threating medical conditions require specialized, comprehensive care and constant monitoring. In middle- and high-income countries many of these patients are treated in an intensive care unit (ICU), where highly specialized nurses and clinicians work together to strive for the patient's survival and optimal recovery. Critical care may be considered one of the most technologically advanced disciplines in modern medicine. Patients in an ICU are continuously monitored, treated with powerful drugs and may undergo multiple organ support. Examples of organ support are the use of mechanical ventilation for respiratory dysfunction, dialysis for renal replacement, and extracorporeal membrane oxygenation for cardiac or pulmonary conditions. This extremely dedicated and advanced care allows patients to survive previously lethal medical conditions such as major trauma, congenital conditions, major surgeries, organ dysfunction or severe sepsis.

ICU mortality has decreased over the years, although it still afflicts 8 to 18% of admitted patients [1]. Decreased mortality shifted the attention towards morbidity and quality of-life after ICU discharge [1]. ICU survivors may face physical and non-physical challenges, such as muscle loss and weakness, chronic pain, depression, long-lasting delirium or neurocognitive impairment. Prevention of the long-term legacy of critical illness and critical care is the main goal of current clinical efforts and research. It is of great interest, on the one hand to identify protective mechanisms against long-term morbidities. On the other hand, to detect the early pathological signs of organ dysfunction or neurological deterioration that may not be directly associated with increased mortality, but could define the patients' quality of life after ICU discharge. Critical care is no longer about survival but about quality of survival.

1.1.2 *Neurological deterioration and neuromonitoring*

Critically ill patients are at risk of long-term neurocognitive impairment, which can result from an initial neurological condition or from secondary complications that affect the brain. Identifying occurring neurological insults and quantifying their effect on brain function can be challenging. First, little information can be gained by clinical examination, considering that most of the ICU patients are unconscious and/or undergo sedation. In addition, more advanced neurological testing requires long-term follow-up.

Second, although neurological insults can be identified through radiological imaging techniques, such examinations are only done on an intermittent basis and require transportation, which can be challenging for the critically ill patient. Third, in pediatric patients the measurement of physiological variables of hemodynamic status of the brain perfusion, which are important contributors to the inadequate substrate delivery to the brain, can be inaccurate [2, 3], making the detection of hypotensive insults challenging. Fourth, neurological damage can originate from many underlying causes, therefore a single monitoring techniques cannot fully describe the complexity of the pathophysiological status of the brain. In the last years, several techniques have evolved to monitor different aspects of the brain status. For example, the partial pressure of brain tissue oxygen ($PbtO_2$) or the near-infrared spectroscopy (NIRS) monitor brain oxygenation. Intracranial pressure (ICP) can be monitored in trauma and non-trauma patients. Microdialysis can be used to monitor regional brain metabolism, while transcranial Doppler (TCD) can measure blood flow velocity. Finally, electroencephalography (EEG) can be used, among others, to detect electrographic seizures. The combination of these techniques create value and provide useful information to the clinician. Last but most importantly, identifying occurring neurological insults remains challenging because neuromonitoring is not yet a perfect science, and even while the link between specific neurological insults and mortality is clear for most pathologies, the exact association between these insults and long-term neurological outcomes remains, in most of the cases, unknown.

1.2 CLINICAL DATA AND MEDICAL DATASETS

1.2.1 *Information overload*

A massive amount of data is generated in an ICU. For each patient, data are generated from several patient monitors, laboratory test analysis, imaging, reports of clinical evaluations and information on the administered treatment (administration of drugs and fluids, initiation of organ support etc.). On a daily basis, the clinical situation of the patient is described with thousands of data points. These data certainly contain valuable information, however their size and complexity exceed human processing capacities. As a result, valuable pieces of information may remain buried within the data. Clinicians are confronted with this problem, also called *information overload*, on a daily basis.

1.2.2 *Electronic health records*

The last decades have seen the progressive digitalization of the ICUs. More advanced technologies have been introduced not only to display information in a more interpretable fashion, but also to store patient data in a more organized way. This led to the creation of electronic health records (EHRs). EHRs are large databases that organize patient data from different sources and different formats, in a structured way. Most of the data are stored into an EHR automatically, with two major benefits. First, EHRs allow the collection of data that previously could not be collected manually, for instance high-frequency monitoring data (mean arterial blood pressure, heart rate etc.). Second, EHRs allow the collection of high-quality medical data, where human error is minimized.

The digitalization of clinical datasets accelerated clinical research, not only by allowing easier and faster access to patient data, but also by facilitating the collection of high-quality multi-center datasets, where data from different centers are collected in a uniform way and then merged.

At the University Hospitals Leuven (UZ Leuven), patient data are collected in the Patient Data Management System (PDMS, Metavision®, iMD-Soft®, Needham, MA, USA). The PDMS collects demographic data, severity scores indexes on admission and daily, continuous monitoring data on a minute-by-minute basis, laboratory test results, reports of clinical evaluation, treatment information and long-term outcomes, for a total of approximately 250 GB of stored data per year (ICU data only).

1.3 BIG DATA AND ARTIFICIAL INTELLIGENCE IN THE ICU

1.3.1 *Overview*

Big Data refers to large and highly complex datasets that require advanced storage methods and powerful processing techniques. In the early 2000s, the analyst Doug Laney defined Big Data with “the three Vs”: Volume, Velocity and Variety. “Volume” describes the enormous size of Big Data dataset. “Velocity” refers to the speed through which data are generated. “Variety” indicates the multitude of different formats (numeric data, text documents, images, etc.) that compose a single Big Data dataset. Two additional “Vs” were later added to the definition, “Veracity” refers to data quality, which depending on the sources of data, may not always be ensured

and finally “Value”, which is intrinsic of data, but needs to be unlocked through advanced analytics.

The widespread implementation of EHRs has led to the creation of huge datasets of clinical data. Although these datasets are relatively small when compared to other industries, we can refer to them as “Big Data”, given that they aggregate information of different formats from different sources. These big clinical datasets are more than just the by-product of patient care, but represent a cheap source of new knowledge [4], which can help reshape and revolutionize several aspects of healthcare, resulting in, for instance, improved prognostication, development of new diagnostic tools or personalized patient treatment [5, 6].

Creating new knowledge from Big Data, often referred to as data mining (DM), may require more advanced methods than traditional statistics. DM includes different algorithms, such as linear regression analysis, Bayesian methods, or more advanced machine learning techniques. Machine learning (ML) refers to a class of algorithms that “learn” how to best perform a task, by recognizing patterns in data and then applying the learned pattern structure to new, unseen data. Through DM techniques it is possible to uncover interactions between variables, find hidden associations, visualize data in an interpretable way or obtain predictions of outcomes, future physiological values or impending threatening events. Data mining techniques can be also more broadly referred to as Artificial Intelligence (AI). AI can be defined as the science of enabling machines to have problem-solving and decision-making capabilities.

Several successful examples on the application of AI techniques on large ICU datasets have been presented. For example, the ML-based prognostic model “DeepSOFA”[7] was demonstrated to outperform the commonly used Sepsis-related Organ Failure Assessment (SOFA) score in assessing severity of illness and predicting in-hospital mortality. Also, it was demonstrated that a ML model can achieve similar discriminative performance as clinicians in the prediction of acute kidney injury in critically ill patients [8]. Similarly, several ML model have shown it is possible to accurately predict, among others, hospital and ICU readmission [9–11], delirium [12] , sepsis [13, 14] or impending brain-threatening events [15].

1.3.2 *Translation to bedside*

Despite the many recent published models, only a very limited number of AI applications are currently used in ICU clinical practice. A recent systematic review by Fleuren et al. [16] has scored 172 published studies that deployed ML models on ICU datasets on their technology readiness level (TRL), which is a scale from 1 to 9, where 9 represents a working technology. Only 5 % of the screened studies presented validated results (TRL = 5), 1% of the studies prospectively validated the model at the bedside in a blinded setting (TRL = 6) and 1% of the articles evaluated the model against clinical relevant outcomes (TRL = 8). Not a single study satisfied the criteria for a TRL of 9.

The lack of AI applications at the bedside can be attributed to several factors. As every new discipline, AI and in particular ML needs to gain trust among clinicians, this is only possible if we strive for transparent and effective reporting, while encouraging replicability [17, 18]. To this end, the new CONSORT-AI and SPIRIT-AI extensions [19, 20] aim at providing new guidelines for the transparent and rigorous reporting of clinical trials that involve the use of ML. In addition, to favor transparency, the TRIPOD guidelines provide recommendation on reporting of any prediction model study regardless of the study type [21]. Another factor that contributes to clinician's trust is model's interpretability, or rather the capacity of the model to provide insights on which input features most contributed to the decision-making process, and to which extent. Interpretability gained incredible value in the last years, with many algorithms that have been developed for that purpose [22]. The level of required interpretability to consider a model trustworthy will mostly depend on center-specific policies and on the result of patient consultation. Beside trustworthiness, these algorithms must be useful. Surprisingly, several published models do not address a clinically meaningful question [17]. Hence, the close collaboration between researchers and healthcare workers is essential to deliver algorithms that effectively respond to the needs of the clinic. Another relevant limitation of AI algorithms for use in clinical practice is that they are often developed on a specific population, but they do not necessarily capture the real world heterogeneity [17, 18]. As a result, the algorithm could be inaccurate when applied to certain subgroups. To these limitations, must be added the ethical and privacy concerns raised by the use of AI models at the bedside, as well as the large economical investment required for the digitization of the ICU.

1.4 DEVELOPMENT OF MACHINE LEARNING MODELS FOR CLINICAL USE

Creating knowledge from Big Data is a delicate process and requires the use of several techniques that range from feature extraction and selection to model development and evaluation of the performance. Throughout the entire process of knowledge creation, a close collaboration between clinicians and data mining professionals is crucial to avoid data misuse or misinterpretation.

The current section aims at providing a general overview of the data mining techniques used in this thesis. As in many new technological fields, data mining uses specific terminology that is summarized in Table 1.1.

1.4.1 *Data quality*

Data collected at the bedside might be prone to human errors, artefacts and missing values. In addition, they are inevitably biased by the standard clinical practice. Hence, a solid knowledge of the underling pathophysiological mechanisms as well as the standard clinical procedures are essential for proper data interpretation. In addition, it is important to perform a careful and objective data-quality assessment before engaging in the analysis, since poor data quality can lead to misleading results. Data-quality can be assessed by computing the number of missing values for specific variables, by checking whether missing values are not linked to the presence of a specific medical condition (which could later bias the results), by performing a visual examination of data or by analyzing the main summary statistics of the data (mean, median, standard deviation, max, min).

Table 1.1 Data mining terminology

TERM	DEFINITION
Big Data	High-volume datasets that collect multimodal and unstructured data from different sources.
Bootstrapping	Method of validation that retrains the model multiple times on randomly selected subsets of the dataset. The final performance of the model is obtained by averaging the performance metrics obtained in every repetition.
Calibration	Calibration quantifies the fit between the probability distribution of the predicted events and the proportion of events observed in a real sample. The better these two match, the better calibrated the model is. Calibration is usually visualized through calibration plots.
Data mining	Set of methods used to create new knowledge from datasets of variable dimension. Data mining can be performed, among others, with linear regression, Bayesian analysis or machine learning techniques.
Development set	Synonym of training set, see below.
Validation set	Dataset on which the model is validated through internal or external validation.
Features	Data elements used to train the model. They ideally contain information that are useful for the goal of the model. For example: demographic characteristics, results of the blood gas analysis or metrics extracted from continuously monitored physiological signals. They are also referred to as predictors, risk indicators or independent variables.
Feature engineering	Process of using statistical methods or domain knowledge to extract meaningful features from a dataset.
Intelligibility	In a model, intelligibility refers to the capacity of understanding the internal mechanisms of an algorithm and identifying the input features that provide the most important information to perform the task.
Machine learning	Set of methods that learn from the data the parameters of the algorithm that allow to perform the required task with the smallest error.
Outcome label	The label is the ground truth or the ideal expected result. The outcome label is compared with the result of the algorithm to evaluate the algorithm's performance. It can be a categorical, ordinal or continuous value.
Overfitting	In machine learning, overfitting indicates the loss of generalizability of the algorithm. The algorithm learns from the particular characteristics of the training data, also called noise. Such particularities are not present in other datasets thus

	leading to loss of performance of the algorithm when applied to unseen data.
Sample	Subset of a population, typically a small fraction. In a dataset that is composed by *np* patients, with *nf* features per patient, a sample is a subset of the dataset with a number of patients < *np*, but same *nf*. A sample can be composed also by 1 observation, e.g. 1 patient with the corresponding *nf* features.
Test set	In machine learning, it indicates the subset of data on which the trained model is tested. The test set must be completely independent from the training set, hence it contains data from different patients or hospitals.
Training set	In machine learning, it indicates the subset of data on which the model is built or "trained".
Validation	Process of evaluating the trained algorithm on an unseen dataset in order to test the algorithm's accuracy, hence its generalizability capacities.

1.4.2 *Feature extraction*

For the sake of simplicity in this section I will use the terms "feature" and "variable" interchangeably, despite the fact that "feature" is a more appropriate term in the context of ML techniques (such as random forests or Gaussian process regressors), while "variable" is more appropriate in the context of classical statistical techniques (such as linear regression, or Bayesian methods).

The accuracy and performance of the chosen technique greatly depends on the quality, quantity and clinical value of the input features. Depending on the research question, input features can be characteristics of the patient, such as age, sex, or a severity of illness score, or they can be extracted from the data in a process called "feature engineering", or both. In feature engineering, mathematical methods are used to extract descriptive statistics from a physiological signal or a set of clinical data. For example, an engineered feature could be the median value of the minute-by-minute MAP signal during the first 24 hours of ICU stay. In this phase, domain knowledge is key to identify the most clinically relevant input features and/or the data from which to extract them.

1.4.3 *Feature selection*

Feature selection is the process of selecting the most important features from a dataset. This phase has important implications on the performance of the model. Feature selection can increase the statistical power of the analysis and reduce the risk that the results exclusively depend on peculiar characteristics of the dataset in use, i.e. overfitting. During feature selection, irrelevant features that can add noise, and therefore negatively affect the model are removed. In addition, the use of a small set of features favors a potential future clinical implementation. On the contrary, a bad feature selection might exclude input features that are actually relevant, loosing valuable insights on the research question.

Stepwise feature selection methods aim at obtaining a subset of the most important features to input to a model, by including or excluding one feature at the time. In this case, feature importance can be estimated with different methodologies such as likelihood ratio or logistic regression models. **Forward selection** starts from an empty subset of features. At every step, feature importance is calculated, and the most important feature is added to the subset of important features. On the contrary, **backward selection** starts from the full set of features and after computing feature importance eliminates at each step the least important feature. A backward selection approach is normally preferred, given that the method takes into consideration the effect of all features simultaneously [23]. The stopping rule for inclusion or exclusion of features is a central issue in stepwise selection methods, the preferred methods are based on the degree to which the model fits the data, while penalizing model complexity.

A more advanced method for feature selection is the **LASSO regression** [24]. When used for feature selection, the LASSO regression is a linear model that minimizes the log likelihood cost function by reducing the number of input features. The reduction of features is performed by shrinking the features' coefficients, depending on which set of coefficients maximizes the cost function. A feature with a coefficient of zero can be discarded. LASSO regression has demonstrated superior results as compared to the traditional feature selection methods [25].

Feature selection can also be performed via **mutual information** [26, 27]**.** Mutual information is a measure of entropy that quantifies the dependency between two random variables. In the context of feature

selection, it measures the reduction of entropy, i.e. uncertainty, for one variable given a known value of a second variable. Features that have high mutual information with the target output compose the subset of selected features. The advantage of this feature selection method is that it captures the non-linear relationship between variables.

In this thesis, the LASSO and mutual information techniques were used for feature selection, the combination of these two methods allows to capture features that have linear (LASSO) and non-linear (mutual information) relationships with the outcome. An example of feature selection from a multidimensional dataset is visualized in Figure 1.1.

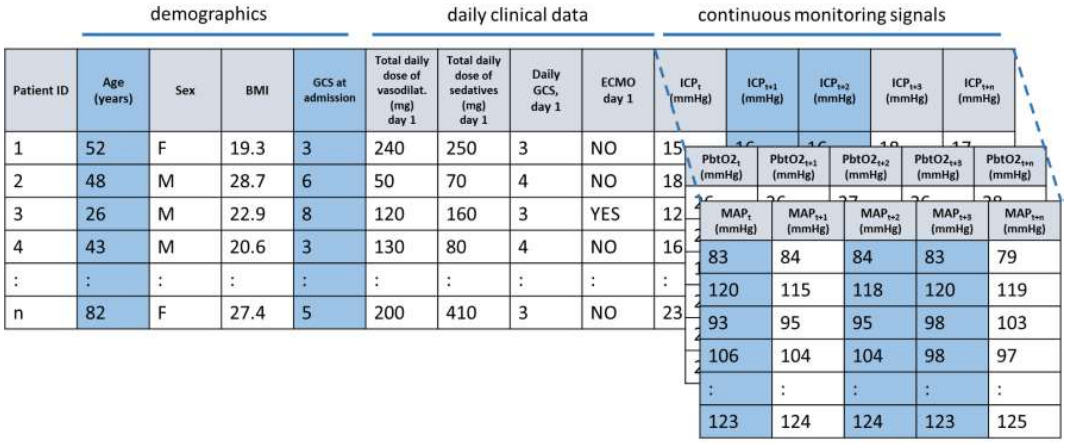


Figure 1.1 Example of feature selection. The feature selection algorithms selects only those features that are relevant to perform the task of interest, represented with the blue columns. Features can be demographic data, daily clinical data, or time-points or features extracted from continuous monitoring signals.

1.4.4 *Data mining techniques*

Four main aspects should be considered when choosing which data mining technique to use: the goal of the analysis, the nature and quality of the input data, the complexity of the problem to tackle and the amount of data available. The task of interest (may be classification, regression or clustering) determines the sub-class of data mining algorithms suitable for the scope. Moreover, depending on the type of input data, the use of some algorithms might present more advantages than others. The complexity of the algorithm should be proportionate to both the complexity of the problem and the amount of data available, although increased complexity

does not necessarily result in better models. Complex algorithms may allow unraveling intricate problems, at the cost of limited understanding of its underlying mechanisms and an increasing demand for bigger training sets. Hence, simpler algorithms should be favored, when possible, to promote intelligibility and to avoid overfitting. In addition, algorithms that use clinically interpretable variables may have a higher probability of being accepted by clinicians and incorporated in clinical practice. This section provides an overview of the data mining techniques used in this thesis.

LOGISTIC REGRESSION Logistic regression is one of the most used statistical models for the analysis of the linear association between a binary outcome and multiple input features. Specifically, the output y of the model is restricted between the range $y \in [0\text{-}1]$ and indicates the probability of an event $p(y=1)$, given the value of the input features. To restrict predictions to the interval [0,1] the model is defined as a linear function in the logistic transformation $logit(p(y=1))$:

$$Logit\big(p(y=1|x_i)\big) = b_i \cdot x_i + a = lp$$

Where a is an estimate for the intercept $\propto$ and b_i is the estimated regression coefficient for β_i.
The coefficients b_i should be interpreted as the effect of a 1-unit increase of the feature x_i on the logit probability of an event. The exponent of the regression coefficients e^{b_i} allows to calculate the β_i coefficients, which indicates the odds ratios. Predicted probabilities can be calculated as: $p(y = 1) = \frac{1}{1+e^{-lp}}$, see Figure 1.2 for the logistic functions. Low logit values indicates low probabilities of an event, for instance a $logit(p(y=1)) = -4$ corresponds to a $p(y=1) = 2\%$.

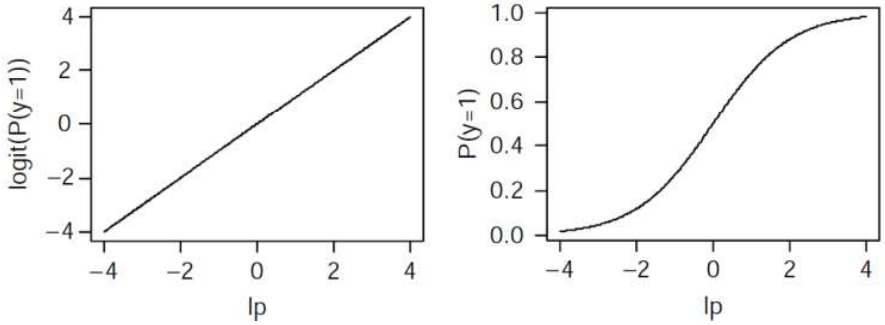


Figure 1.2 Logistic function

BAYESIAN LINEAR REGRESSION Bayesian statistics uses probabilities to expresses knowledge about an unknown parameter. To explain the main principles of Bayesian analysis I will start by giving a brief introduction of traditional linear regression (frequentist approach). In traditional linear regression, the outcome y is expressed as a linear combination of the x_i variables with the estimated regression coefficients β_i.

$$y = \alpha + \beta_i \cdot x_i + error$$

The β_i coefficients indicate the increase in y corresponding to a 1-unit increase of x_i and are often expressed with an absolute number, a confidence interval and a p-value.

Bayesian statistics describes the regression coefficients β_i using probability distributions rather than point estimates, see Figure 1.3. As a result, the effect that x_i has on the outcome y is expressed in terms of probabilities. The use of probability distributions to represent β_i allows to model the uncertainty derived from trying to capture general associations by analyzing a limited number of observations.

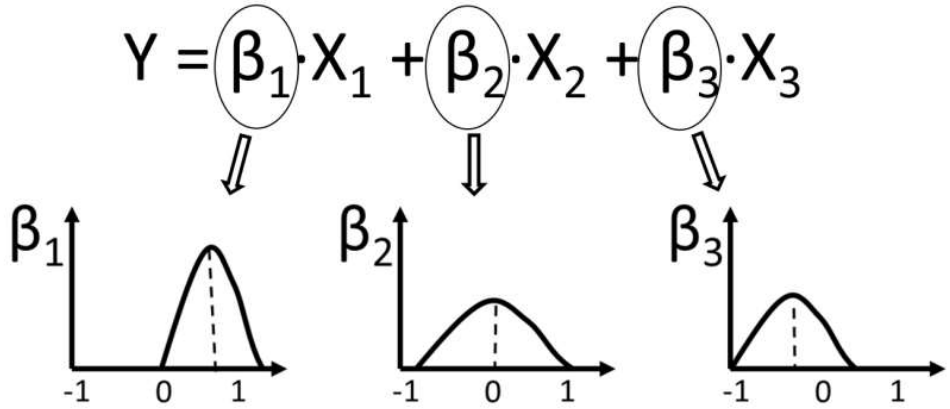


Figure 1.3 Example of the conceptualization of a linear regression problem as formulated in Bayesian statistic. The regression coefficients β_i are expressed in terms of probability distributions. For example, the β_1 regression coefficient of the variable X_1 has a positive mean value ($\mu_{\beta 1} > 0$) and an entirely positive probability distribution. This means that an increase in X_1 increases the probability of higher y, where $\mu_{\beta 1}$ represents the most likely increase in y given a 1-unit increase of X_1. The β_2 regression coefficient of the variable X_2 has a mean value equal to zero ($\mu_{\beta 2} = 0$), therefore it is likely that X_2 is uninfluential on the output value. The β_3 regression coefficient of the variable X_3 has a negative mean value ($\mu_{\beta 3} < 0$) and a partially negative probability distribution. This means that there is a high chance that an

increase in X_3 will result in a decrease of y, however there is also a small chance that an increase in X_3 will result in an increase of y.

In Bayesian analysis, the β_i of the linear regression fitted model (posterior β estimates), integrate prior knowledge (results of previous studies or common knowledge) with the information provided by the analyzed data, see Figure 1.4. Prior knowledge is provided through the "priors", which, in this case, represent the predictions about the probability distribution of the β estimates before additional information (the analyzed data) becomes available. For example, we want to study the association between neurocognitive outcomes and duration of hypoxia in children that undergo cardiac surgery. We know that other factors, such as the presence of severe genetic disorders will affect the outcomes of the children irrespective of hypoxia. This is "prior knowledge" or, in other words, validated information that can be added to the model even before the model is applied to the data. Bayesian statistics allows to include this knowledge in the model, by setting the "priors", or rather the expected probability distribution for a given β estimate. Coming back to the previous example, we might set for the β estimate of the variable "presence of genetic disorder" a negative probability distribution, to indicate that even before seeing the data our prior knowledge is that the presence of a severe genetic disorder will have a negative effect on the neurocognitive outcomes. In the same analysis we might want to correct for other factors, for example the age of the patient, of which we do not know the effect on outcomes. In that case we can set a "non-informative prior", which is a flat probability distribution where every value of β is equally likely. The inclusion of prior knowledge is important in analyses with small-size datasets, where outliers have a bigger influence on the results.

The results of the Bayesian models are presented as the most likely values of the posterior β estimates and credible interval. The credible interval (C_RI) of the posterior distributions indicates the range of values to which the β estimate belongs with a certain probability.

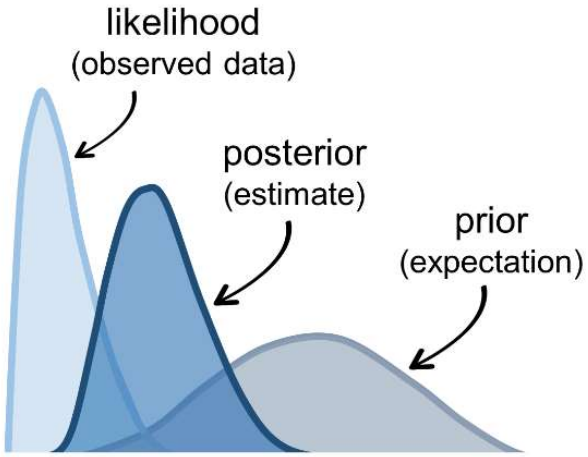


Figure 1.4 Visual representation of the contribution of a prior distribution (expectation on the relationship between a certain variable and the outcome of interest) and a likelihood distribution (relationship between a certain variable and outcomes of interest as observed on the data in analysis) on the posterior distribution (estimate of the relationship between a certain variable and outcomes of interest in the general population).

GAUSSIAN PROCESSES Gaussian processes (GP) have been successfully used in non-linear regression problems because of their flexible modeling abilities [28]. A Gaussian process is a generalization of the Gaussian probability distribution. Whereas a Gaussian probability distribution describes random variables, a Gaussian process is a probability distribution over functions [28]. To make an example, given a set of data points there are potentially infinite functions that could fit the data. A Gaussian process assigns a probability to each of these functions. The mean of this probability distribution then represents the most probable characterization of the data. A visual example of Gaussian processes is given in Figure 1.5. In more rigorous terms, the Gaussian Process is specified by a mean function $\mu(x)$ and a covariance function. The mean function $\mu(x)$ describes the expected value of the function across each dimension of $\mu(x)$. The covariance function models the variance along each dimension and determines how the different random variables are correlated. It can be calculated by applying a covariance function $\mathrm{k}(x_i, x_j)$, or *kernel*, to the data. The *kernel* determines the main properties of the function, for instance, whether it describes a periodic or linear trend or a smooth or lumpy function. Therefore a Gaussian process can be defined as:

$$f(x) \sim GP(\mu(x), \mathrm{k}(x_i, x_j))$$

An example of kernel function is the Rational quadratic kernel, which is given by:

$$k(x_i, x_j) = \sigma^2 \left(1 + \frac{d(x_i, x_j)^2}{2\alpha l^2}\right)^{-\alpha}$$

Where σ^2 is the overall variance and l is the length-of-scale. Increasing values of σ^2 result in a bigger difference in $f(x)$ between adjacent points, while increasing values of l make points that are far-away more correlated. If points x_i and x_j are similar by $k(x_i, x_j)$ the function value at these points ($f(x_i)$ and $f(x_j)$) is expected to be similar.

I will further consider an example without noise functions. Once that $\mu(x)$ and $k(x_i, x_j)$ are specified, the GP training set (GP prior) consists of a distribution of function values f at inputs X. After the training, the GP prior is converted into a posterior $p(f_*|X_*, X, f)$. The posterior distribution can be used to make predictions f_* at new inputs X_*. The joint distribution of observed values f and the predictions f_* can be written as:

$$\begin{pmatrix} f \\ f_* \end{pmatrix} \sim N\left(0, \begin{pmatrix} K & K_* \\ K_*^T & K_{**} \end{pmatrix}\right)$$

Where $K_* = k(X, X_*)$ and $K_{**} = k(X_*, X_*)$.

The predictive distribution can be expressed as:

$$p(f_*|X_*, X, f) = N(f_*|\mu_*, \Sigma_*)$$

$$\mu_* = K_*^T K^{-1} f$$
$$\Sigma_* = K_{**} - K_*^T K^{-1} K_*$$

The resolution of this equation requires additional mathematical steps that go beyond the purpose of this thesis and therefore will not be discussed. It is however important to know that fitting a GP model requires complex mathematical transformations that most of the times can be solved only by approximation, for this purpose the Cholesky decomposition is often used.

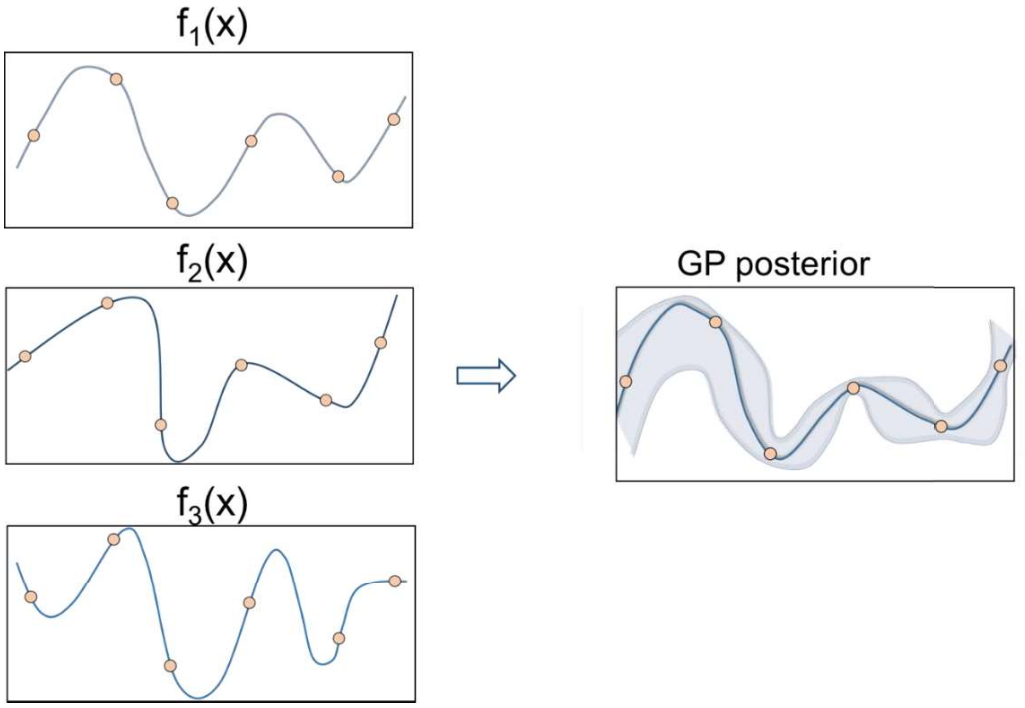


Figure 1.5 Visual example of Gaussian Processes. On the left, three examples of potential functions that fit the same set of data-points. On the right, a visual example of Gaussian process posterior for the same set of data points. The line in blue represents the mean value of the probability distribution, which is also the function that most likely fits the data. The area in light-blue represents the probability distribution over functions that fit the data points.

DECISION TREES Decision trees (DTs) are ML models which can be used for regression or classification [29]. They are built by recursively partitioning the data points of a dataset in sub-groups. A specific output, for example a class or a discrete value, is assigned at each of the final obtained sub-groups. Hence a DT can be seen as a series of "questions" that allow to achieve a desired task with the best performance. The partition is based on the maximization of a cost function and it proceeds until the combination of questions that best fits the dataset is found or until a stopping condition is reached. The series of partitions that compose a DT can be represented graphically in a flowchart-like structure, see example of Figure 1.6.

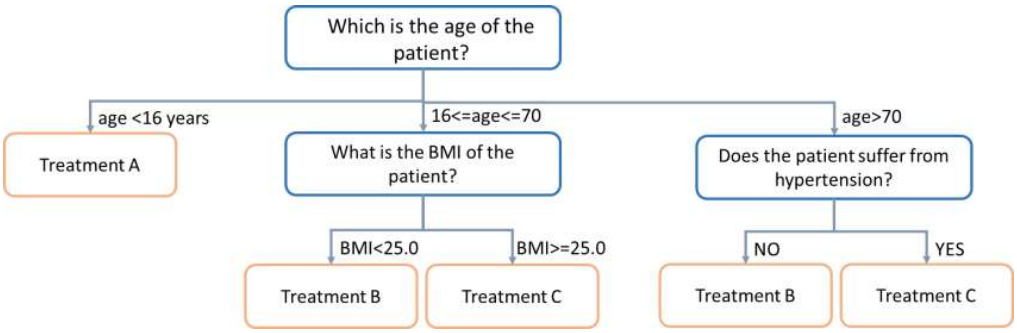


Figure 1.6 Example of decision tree for a clinical problem. In the simplified example, the decision tree is used to decide which treatment is better for a patient based on patients' characteristics. BMI = body mass index.

DTs have several advantages, first they can capture linear and non-linear associations. Second, they do not require big amount of data to be trained. Third, they are interpretable. Interpretability is an important aspect when performing research on clinical data. However, a major disadvantage of DTs is their risk of overfitting, which occurs when the questions used for the partitions are not generalizable enough to other unseen datasets. Also, DTs can be unstable, in fact small variations in the data can lead to a completely different tree being generated. To avoid overfitting and instability, it is recommended to limit the depth of the trees, or rather the number of successive splitting questions that can be used to achieve a task. Another measure to reduce overfitting is to set a minimum number of samples that are required to further split the data.

RANDOM FOREST Random forest (RF) is an ensemble ML model that combines multiple DTs [30], see Figure 1.7. Each DT is built on a random selection of samples from the dataset (selection with replacement, also called bootstrap aggregation), the final output of the RF model is a weighted mean of the outputs of each DT that composes the model. This methodology improves performance as compared to a simple DT and reduces the risk of overfitting and instability. The success of RF relies on the fact that each DT that composes the RF model is quite different from the others, and as such, each DT captures different patterns in the data. Low correlation between DTs is obtained by training each DT on a randomly selected subset of data and on a randomly selected subset of features. An additional optimization

parameter to avoid overfitting is the maximum number of trees that can compose a RF model.

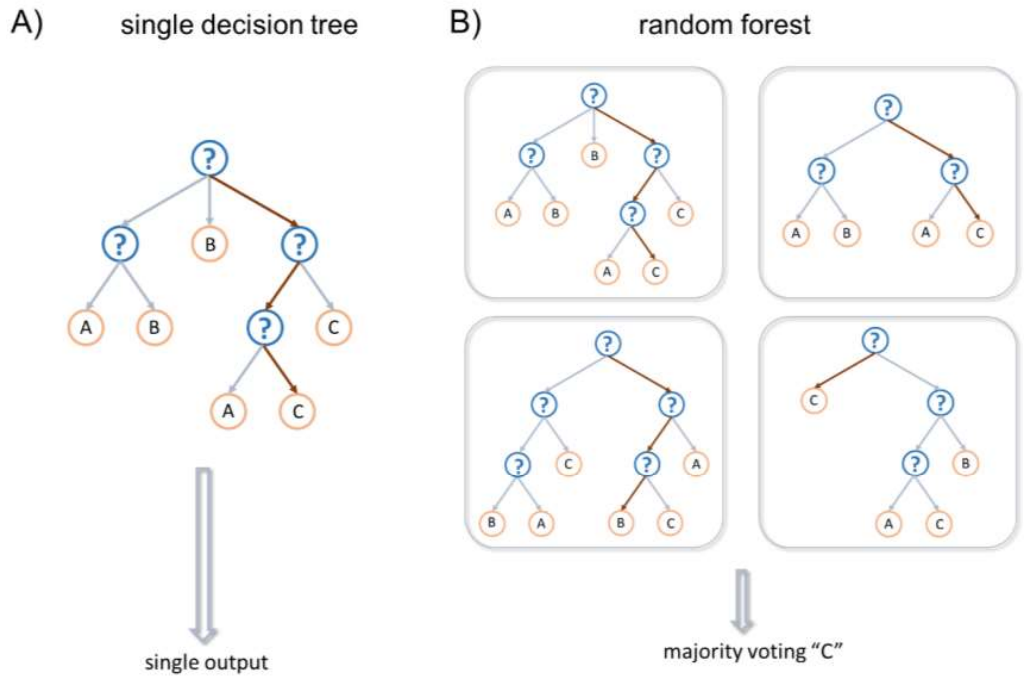


Figure 1.7 Comparison between decision trees and random forest. The possible output of the model are options "A","B" or "C". The splitting questions of each decision trees are represented with a blue question mark. Panel A) decision tree. The tree has a single output, in this example the model outputs option "C". Panel B) random forest. In this example, the random forests model is composed by 4 decision trees, the majority of the trees votes option "C", therefore "C" will be the output of the random forests model. Each of the decision tree that compose the random forest is trained on a randomly selected subset of samples and features.

1.4.5 *Evaluation metrics*

Model performance refers to the degree of agreement between the results of a model and the ground truth. Performance evaluation has many facets, and as such, it is evaluated through different statistical measures that provide complementary information.

In the case of a classifier, true positives (TP) and true negatives (TN) indicate respectively the number of positives and negative samples that are correctly classified. False positives (FP) and false negatives (FN) indicate

respectively the number of negative and positive samples that are misclassified. Normally the condition of interest is labelled as positive.

The discrimination of the model, or rather the capacity of the model to correctly discriminate between patients with and without a condition of interest can be quantified with different metrics such as: accuracy, precision, negative predictive value, sensitivity and specificity.

Accuracy indicates the overall ability of the model to discriminate between classes.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision, or positive predictive value (PPV), is a measure of accuracy that is exclusively limited to the condition of interest. It indicates the probability of having the condition of interest after a positive test result.

$$precision = \frac{TP}{TP + FP}$$

Negative predictive value (NPV), indicates the probability of not having the condition of interest, after a negative test result.

$$NPV = \frac{TN}{TN + FN}$$

Sensitivity, or recall, indicates the ability of the model to correctly identify the condition of interest.

$$sensitivity = \frac{TP}{TP + FN}$$

Specificity indicates the ability of the model to correctly identify when the condition of interest is not present.

$$specificity = \frac{TN}{TN + FP}$$

The **receiver operating characteristic curve (ROC curve)** is a plot that displays the performance of a classification model in terms of sensitivity and 1 – specificity at all classification thresholds. Through the ROC curve it is possible to appreciate the trade-off between sensitivity and

specificity. In fact, lowering the classification threshold results in increased sensitivity at the cost of lower specificity, and vice versa. To maximize performance, the optimal classification threshold can be computed from the ROC by extracting the threshold on the curve that is the closest to the top-left corner of the plot. The area under the ROC (AUROC or AUC) is an aggregated measure of the performance of the model across the entire set of classification thresholds. AUC can range between 0.5 and 1, where 0.5 would indicate a random classifier, and 1 a perfect model. Depending on the problem, an AUC of 0.7-0.8 is considered adequate for clinical models [23]. The advantages of the ROC and the AUC is that they are scale-invariant, therefore their measure does not depend on the absolute value of the predictions but rather depends on the predictions' rank. In addition, the ROC and the AUC give an objective estimation of the discrimination abilities of the model that does not depend on the chosen classification-threshold. However, these two metrics have also some limitations. They do not provide any quantification of calibration, and they may not fully describe the model discrimination capacities when high sensitivity or high-specificity are required.

Differently from the ROC curve the **precision recall curve** displays the precision of a classifier as a function of recall (also called sensitivity) at all classification thresholds [31]. The trade-off between precision and recall can be useful in selecting an optimal threshold. The **area under the precision recall curve (AP)** measures the performance of the model across the entire range of thresholds. The optimal threshold can be extracted from the curve, and it corresponds to the point that maximizes the harmonic mean of precision and recall;
*f = (2*precision*recall)/(precision+recall)*. Differently from the AUC, the AP does not depend on the prevalence of the condition of interest in the dataset in analysis. AP ranges between 0 and 1, with random performance equal to sample prevalence in the local dataset.

Calibration refers to the degree of agreement between the incidence of a condition of interest in a population and the incidence of the same condition in the predictions of the model, see example in Figure 1.8. For instance, in a population of 100 patients where the incidence of asthma is 20%, a well-calibrated model will predict this condition on about 20 patients. Model calibration can be visually assessed with the calibration curve, which plots the predicted probabilities of a condition of interest versus the observed proportions in the population. In case of a binary

output, the continuous range of observed proportion is calculated across bins with the Loess algorithm [32]. Normally the calibration curve is compared with a diagonal line at 45° that represents a perfectly calibrated model, lack of statistical difference (p-value>0.05) with this line indicates good calibration. The calibration curve can be characterized with two statistics: the calibration slope and the calibration-in-the-large. The first one indicates the slope of the fitted curve, while the latter indicates the intercept of the fitted curve on the y-axis. A perfectly calibrated model has calibration slope = 1 and calibration-in-the-large = 0. The degree of uncertainty around the calibration curve is plotted with the calibration belt, which indicates the range of values to which the general calibration curve belongs given a certain confidence level [33].

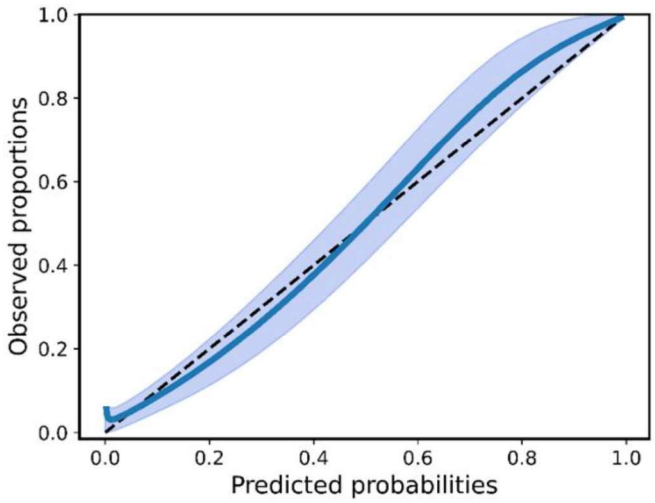


Figure 1.8 Example of a calibration plot. The line in blue indicates the calibration curve, while the area in light-blue indicates the calibration belt for a confidence level of 0.95. The dashed line in black represents a perfectly calibrated model.

Clinical usefulness represents the clinical benefit that derives from the use of a model as compared to the standard policy without the model and it evaluates the performance of a model in a decision-analytic perspective. To determine clinical usefulness, it is important to set a "decision threshold" or "classification cut-off" that triggers the medical intervention. The definition of such threshold does not depend on statistical criteria, but rather on the context of the decision. For example, in the evaluation of a classifier for cancer diagnosis, it may be more relevant from a clinical point

of view to minimize the number of false negatives and therefore avoid misdiagnosis, rather than to avoid false alarms. On the contrary, in the evaluation of a model that provides decision-support for the administration of a drug with very severe side effects, a model that minimizes false positives may be preferred over a very sensitive model. The best decision threshold depends on a trade-off between the harm and benefits that are associated with a classification or a prediction. As a result, the best decision threshold for a model might not necessarily coincide with the threshold that maximizes performance metrics such as accuracy or precision, but it could rather be the threshold that better adapts to the needs of the context in which the model is used.

However, in practice, it is difficult to define an optimal decision threshold precisely. First, to find a threshold that applies to the general population a huge amount of data is needed. Second, optimal decision thresholds may vary across patients, depending on their clinical history and personal needs. In a real-life context, a range of decision thresholds may be more appropriate. The range of decision thresholds in which the model presents clinical usefulness can be visualized with a **decision curve**. A decision curve is a visual representation of the clinical usefulness of a model across the range of possible decision thresholds [34]. In a decision curve clinical usefulness is calculated as net benefit (NB), see equation below.

$$NB = \frac{TP - wFP}{n}$$

Where w indicates the ratio of harm-to-benefit and n is the total number of subjects over which the net benefit is computed.

The NB of a model is compared to the default policies “intervention for all” and “intervention for none”, where “intervention for all” means that every patient is treated for a certain condition irrespective of whether they may have it or not and “intervention for none” means that no patient is treated, and therefore the NB is equal to zero. A model is clinically useful if it has larger NB than the default options “intervention for none” and “intervention for all” for all the relevant decision thresholds, as shown in the example of Figure 1.9.

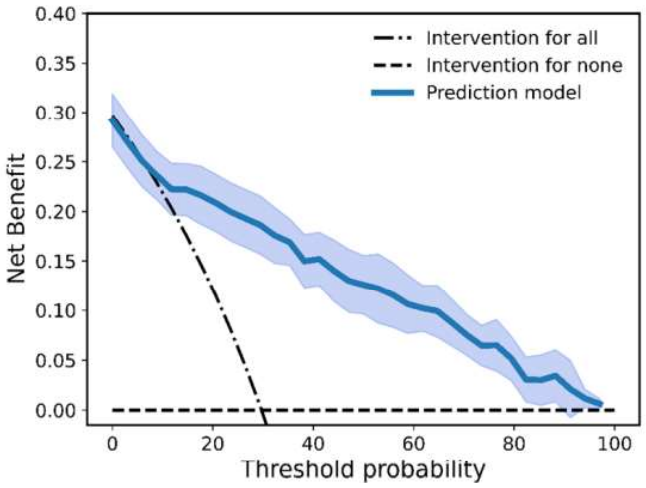


Figure 1.9 Example of decision curve. The curve plots the net benefit versus the corresponding decision threshold. The line in blue indicates the net benefit (NB) of the model, the light-blue area indicates the 95% confidence interval of the curve. The dashed line indicates the net benefit of the default treatment "intervention for none", while the dashed-dotted line indicates the NB for the default treatment "intervention for none".

1.4.6 *Validation techniques*

Validation is an essential step in data mining and refers to the assessment of the model performance on previously unseen data. Two types of validations can be performed. **Internal validation** is a minimum prerequisite of model development, and it assesses the performance of the model on a previously unseen dataset that is generated from the same setting as the development dataset. During model development, internal validation can be used to avoid overfitting and reduce bias in model performance. **External validation** assesses the performance of the model when applied on a completely independent dataset, and as such, it measures the capacity of the model to generalize.

One technique for internal validation is **k-folds cross-validation**, where the development cohort is randomly divided into *k* subgroups (or folds). Iteratively, the model is trained on *k-1* folds, and internally validated on the remaining fold, see the example at Figure 1.10. The procedure is repeated until all 10 folds have been used as internal validation cohort. The final performance of the model is given by the average of the performance on the

k internal validation cohorts. Normally used values for *k* are 3, 5 and 10. Cross-validation has been proven to be a very robust technique for model development, however it is not recommended in case of small datasets, given that the splitting of the data in *k* folds may decrease power.

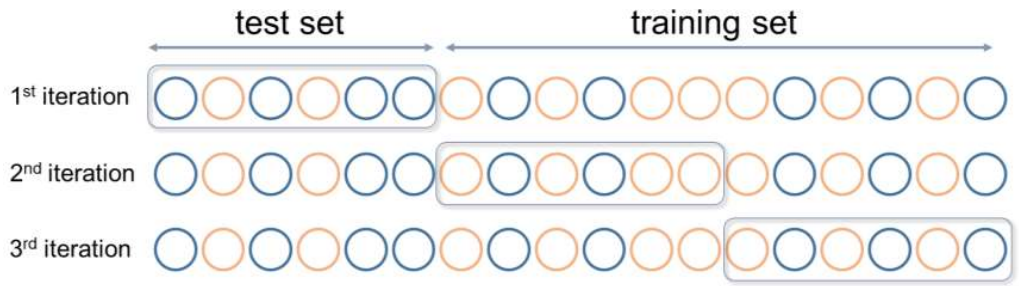


Figure 1.10 Example of k-folds cross validation, where k =3. At every iteration, the dataset is divided into 3 folds, 2 folds are used to train the model and 1 fold for internal validation. Each fold may have a different prevalence of patients with or without the condition of interest, in orange and blue respectively. The final performance of the model is the average of the performance of the model on each testing set.

External validation can be further divided in temporal and/or geographical validation. In temporal validation, there is temporal independence between the development and external validation cohort. For example, two datasets can be considered temporal independent if they are collected in the same setting but 10 years after one another. In geographical validation, development and validation cohorts are collected in two geographically different settings. For example, two datasets simultaneously collected in two different European countries can be considered geographically independent. Therefore, external validation assesses whether the model is robust towards changes in clinical practice or center-wise differences in patient management or data acquisition. Also, external validation may help identifying whether a model needs to be recalibrated if used on a new population. It is important to externally validate the model on sufficiently large datasets, so that lack of statistical power does not result in misleading conclusions about the generalizability of the model.

1.5 RESEARCH QUESTIONS

This thesis contributes to the line of research that aims at reducing the long-term legacy of intensive care, with a focus on neurological outcomes. More specifically, the thesis will focus on three main objectives:

- Unraveling complex associations between neuro-monitoring signals and long-term neurological and neurocognitive outcomes.
- Predicting the occurrence of potentially harmful intracranial events that may lead to poor long-term neurological outcomes.
- Developing and validating a decision support application that brings part of the knowledge acquired in this thesis to the bedside.

Three different types of neuro-monitored patients are studied: pediatric patients after surgery for congenital heart defects, patients with sub-arachnoid hemorrhage and patients with severe traumatic brain injury.

1.5.1 *Pediatric patients after surgery for congenital heart defects*

Congenital heart defect (CHD) is a structural malformation of the heart or great vessels [35]. It includes cardiac defects such as patent ductus arteriosus, atrial septal defects, aortic stenosis or bicuspid aortic valve [35]. This pathology, which affects about 8 out of every 1000 newborns [36, 37], can have enormous repercussions on the cardiac function and systemic blood circulation of the child, to the extent that about 25% of cases require corrective surgery in the first year of life [37].

Although most cases of CHD can be corrected through surgery with low mortality, children with severe CHD often present long-term neurodevelopmental deficits [36]. Reduced neurodevelopment can be linked to several combined factors, but one key aspect is the neurological damage provoked by insufficient cerebral oxygenation, which can occur during the gestational and perioperative period [36].

At birth, children with severe cases of CHD often present gray and white matter injuries or signs of maturation delays [38–40]. White matter lesions are the more extensive, and tend to expand or even appear after birth, during the intraoperative and early post-operative period [41, 42]. The main cause of these lesions are hypoxic and ischemic events [36], which are a consequence of the frequent post-operative fluctuations in cerebral oxygenation, due to limited or absent cerebrovascular autoregulation [36,

43–45]. Therefore, to maximize outcomes, clinical efforts aim at providing optimal cerebral perfusion during the entire perioperative period.

Monitoring cerebral perfusion is challenging in pediatric patients [46]. Invasive hemodynamic monitoring is possible, but not free from complications such as bleeding, arterial obstruction, thrombosis of central venous circulation and sepsis [47, 48]. CT perfusion imaging is non-invasive, but sporadic. Moreover, depending on the hospital organization, the execution of imaging examinations may require long transportation outside the ICU. Transportation poses inherent risks for the critically ill child, due to the physical displacement of the patient, changes of organ-support devices, equipment failure due to transport or dislodgement and disconnection of some monitoring devices. Given the limitations of the traditional methods, in the last years Near-Infrared Spectroscopy (NIRS) has gained importance as a non-invasive method to assess continuous measurements of regional brain tissue oxygen saturation ($SctO_2$) [49, 50]. An example of NIRS monitor is displayed in Figure 1.11.

NIRS is a light-based technology which relies on the capacity of near-infrared light to propagate a few centimeters into biological tissues, including skin and bones [51–53]. The propagation of light in the tissues depends on three main factors: reflection, absorption and scattering [51]. Light absorption is a key element in the functioning of NIRS. In fact, each compound present in the tissue absorbs near-infrared light differently, where the overall attenuation of the incident light is given by the sum of light absorbed by each compound [51]. NIRS technology is based on the fact that the absorbance spectrum of hemoglobin differs depending on whether it is bounded with oxygen (oxygenated vs deoxygenated hemoglobin) [51–53]. The use of near-infrared light at different wavelengths allows for the calculation of the relative concentrations of oxygenated and deoxygenated hemoglobin in relation to the total hemoglobin concentration, allowing for a measure of oxygen saturation [51, 52]. In a pediatric setting the NIRS cerebral oximeter can be used to measure the local cerebral tissue oxygen saturation of the frontal lobe of the child.

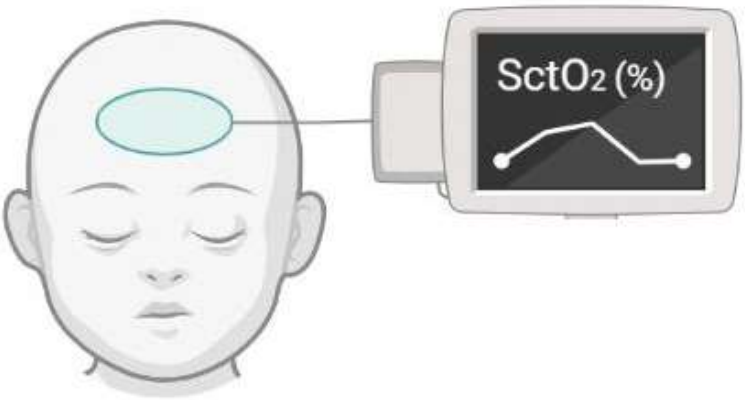


Figure 1.11 Example of near infrared monitoring. For neuro-monitoring purposes, the electrodes are placed on the forehead of the child.

Although NIRS is a promising tool for the neuro-monitoring of children with CHD, the range of safe physiological $SctO_2$ values is not well defined [52], and may lie between 55% and 80%. As a result, clear guidelines on $SctO_2$ interpretation and treatment are lacking, and the common clinical practice is limited to preventing events of extreme hypoxia ($SctO_2$<40-50%) and of hyperoxia ($SctO_2$>85-90%). The potential clinical benefit of this practice remains a topic of discussion [54].

Over the last years, scientific focus has shifted to better identifying the lower end of the safe $SctO_2$ range. Interestingly, recent studies have shown that reduced $SctO_2$ is not only associated with increased mortality but also with increased neurodevelopmental impairments [55, 56], suggesting that neurological damage may already occur at higher $SctO_2$ levels. Therefore, in children undergoing corrective surgery for CHD it remains unclear which are the $SctO_2$ target levels to reduce mortality and to avoid the long-term neurocognitive impact of the peri-operative period. Several studies have tried to address this question but with inconsistent results, mostly due to the limited number of children enrolled.

1.5.2 *Traumatic brain injury*

Traumatic brain injury (TBI) is defined as an alteration of brain function caused by the impact of a strong external force on the head [57]. This traumatic event can occur, among others, because of a traffic or sport accident, or as a result of a fall. TBI is a major cause of death and disability worldwide, with a mortality rate around 15%-30% [58], and almost half of

the survivors that experience long-term physical, psychiatric and cognitive disabilities [57, 58]. It is referred to as a “silent epidemic”, given that it accounts for approximately 2.5 million new cases each year, only in Europe [57].

In TBI, the initial traumatic event triggers a series of pathophysiological mechanisms that can lead to additional brain damage, called secondary brain injuries (SBI). Clinical efforts are not focused on treating the initial injury, which is considered irreversible, but on preventing SBI from occurring. Management of TBI aims at achieving four main targets: optimization of cardiorespiratory physiology, safeguarding nutrient and oxygen supply, maintenance of cerebral perfusion pressure (CPP) and control of intracranial pressure (ICP) [57]. The ICP is the pressure that fluids such as the cerebrospinal fluid (CSF) and the arterial and venous blood, exert on the brain tissue inside the skull and the CPP is the pressure gradient that allows blood to enter the brain. CPP can be calculated as the difference between the mean arterial blood pressure (MAP) and the ICP.

Management of TBI is based on a staircase approach of increasing treatment intensity [59, 60], even though none of these treatments are risk-free and, when applied inappropriately, can be associated with worse outcome. Recent SIBICC recommendations [60] classify as first-tier therapies: sedation, hyperosmotic infusions, limited hyperventilation, and drainage of cerebrospinal fluid. Second tier therapies include more intensive hyperventilation, administration of neuromuscular blockade and in presence of functioning autoregulation increasing CPP with fluid boluses, vasopressors and/or inotropes. The most aggressive tier, also referred to as third tier, includes mild cooling (to core temperatures of 35–36°C), deep sedation for metabolic suppression and decompressive craniectomy (surgical removal of a portion of the brain skull to allow brain swelling). Third tier therapies should be considered as last resorts therapies in case the patient did not respond to previous treatments.

Most of these treatments aim at preventing or controlling extreme ICP elevations, which are one of the main causes of SBI. Elevated ICP has been associated with poor neurological outcomes and death in multiple studies [57, 61, 62]. After the initial traumatic event, systemic and intracranial events such as edema, ischemia, or contusion expansion, can provoke the rise of the ICP to potentially dangerous levels. Elevated ICP can impair the CPP, putting brain perfusion at risk, but also it can mechanically

distort the brain, leading to herniation of the brain steam and death. An example of physiological and pathological ICP and ICP monitoring is provided in Figure 1.12.

ICP monitoring may be performed through two methodologies. In the first methodology a catheter, called external ventricular drain (EVD), is placed in a lateral ventricle and connected to an external transducer through a fluid-filled system. This allows to obtain an ICP measurement, but also to drain cerebrospinal fluid (CSF), if needed. An EVD is considered the standard of care for ICP monitoring, however it may provide inaccurate readings during CSF drainage. Alternatively, a probe may be inserted in the intraparenchymal space of the brain, this is called intraparenchymal ICP monitoring. Intraparenchymal monitoring alone does not allow for CSF drainage but it provides more reliable ICP readings. In the ICU of UZ Leuven intraparenchymal ICP monitoring is considered the gold standard and it is often used in parallel with an EVD catheter.

Although optimal treatment thresholds for elevated ICP remain uncertain, current guidelines suggest to start aggressive treatment (third-tier therapies) when the ICP rises above 22 mmHg [63]. The clinical validity of a threshold based strategy is debated [64, 65]. First, the proposed threshold is population-derived (from a single-center study), therefore it does not allow therapy to be targeted to specific subgroups of patients. Second, some studies have shown that ICP levels lower than 22 mmHg might also be associated with worse neurological outcomes [66]. Third, secondary injury by elevated ICP is not adequately defined by the simple, sometimes brief, crossing of a universal threshold.

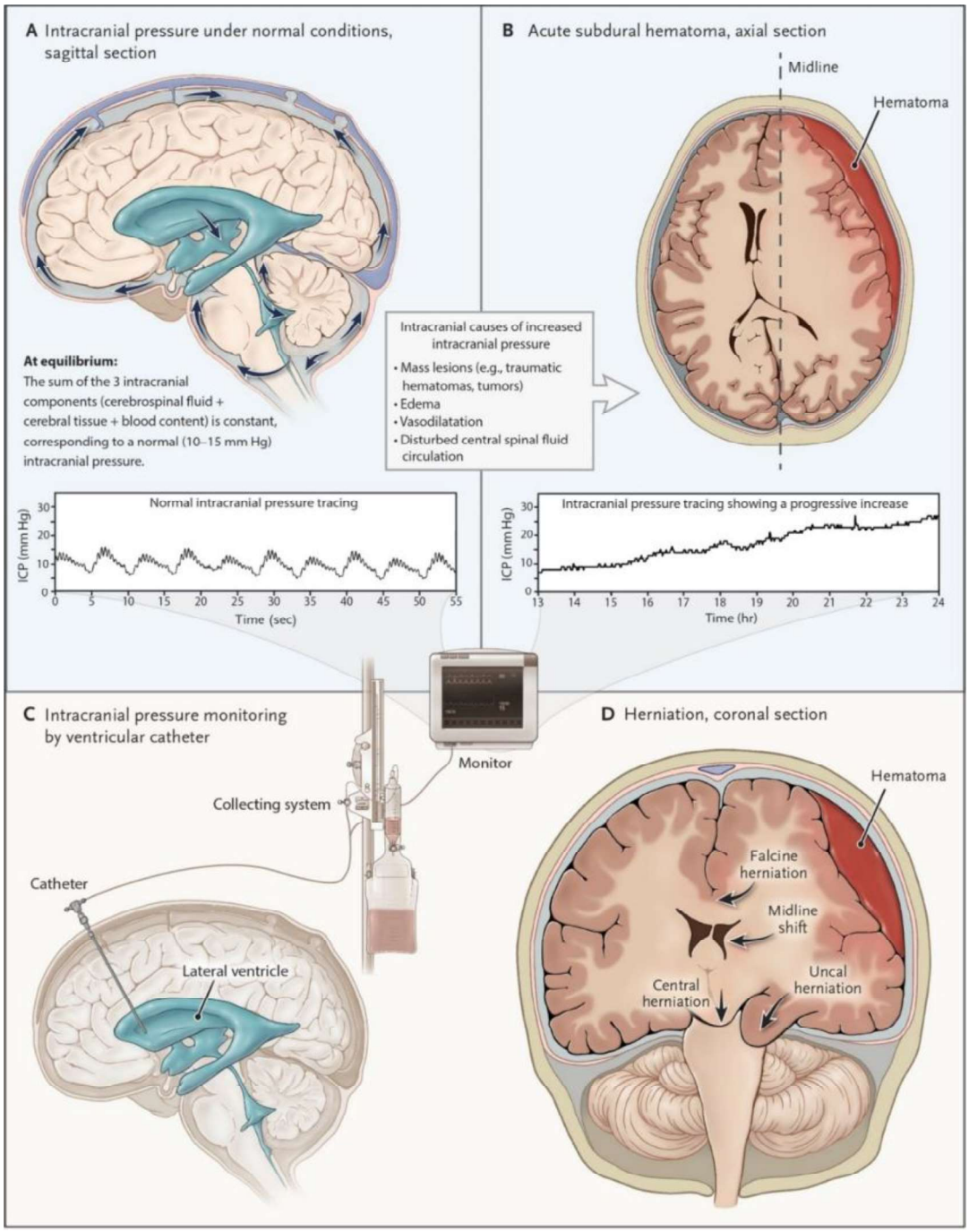


Figure 1.12 Intracranial pressure under normal and abnormal conditions. Under normal conditions, the intracranial pressure (ICP) remains constant at 10 to 15 mm Hg, fluctuating with cardiac and respiratory cycles, as shown in the normal trace recording (Panel A). Since the cranium is a rigid container, the sum of the various intracranial volumes (brain tissue, cerebrospinal fluid, and blood) must remain constant. Cerebrospinal fluid is continuously formed and reabsorbed, with the circulation indicated by blue arrows. Several intracranial and systemic causes may

alter the intracranial components and cause pathologic increases in intracranial pressure. For example, a traumatic left subdural hematoma that compresses the brain and shifts the lateral ventricles to the right, as shown on computed tomography (Panel B). The hematoma volume cannot be compensated by buffering systems, and there is a corresponding increase in intracranial pressure, which can be recorded through a catheter inserted in a lateral ventricle (also allowing the withdrawal of cerebrospinal fluid) (Panel C). The catheter is connected to a collecting system, to which cerebrospinal fluid can be drained, and to a monitor, where the trace recording of intracranial pressure is displayed. Intracranial hypertension may cause compression and displacement of the cerebral tissue from areas of higher pressure toward areas of lower resistance (Panel D). Brain herniation occurs in three main ways. First, a hemisphere is displaced medially against the falx, resulting in falcine herniation. Second, a unilateral pressure gradient pushes the medial edge of the temporal lobe (uncus) through the tentorial foramen, resulting in uncal herniation. In this syndrome, the third cranial nerve and the posterior cerebral artery are compressed, causing unilateral pupillary dilation, a lack of reactivity to light, and infarction. The brain stem is distorted and compressed, with early impairment of consciousness. Third, a bilateral, homogeneous increase in intracranial pressure in the supratentorial space displaces the brain downward through the tentorial foramen, resulting in central transtentorial herniation. The brain stem is compressed and displaced downward without signs of lateralization and with bilateral pupillary abnormalities. Reproduced with permission from N. Stochetti, A. Maas, Traumatic Intracranial Hypertension, N Engl J Med 2014; 370:2121-2130. Copyright Massachusetts Medical Society.

The ICP dose, *i.e.* the combination of intensity and duration of an ICP event, might offer a better representation of SBI due to elevated ICP [67]. The dose is a composite measure that takes into consideration not only the ICP intensity but also its duration, as represented with the gray area in Figure 1.13. Elevated ICP doses have been associated with poor long-term neurological outcomes [66–71], indicating that changes in ICP over time may have a greater clinical relevance than ICP as an absolute value. The association between doses of ICP and worse outcomes, can be visualized with the plots introduced by Güiza et al. [66], see Figure 1.14, that was further replicated in other large datasets [68, 72]. An exponential black line, also called transition curve, separates the ICP doses that occur more frequently in patients with worse neurological outcomes (represented in red) from the ICP doses that occur more frequently in patients with better outcomes (represented in blue). These studies suggest that high ICP values can be tolerated if maintained for a short period. On the contrary, ICP

values between 15 mmHg and 22 mmHg, if maintained for a prolonged time, could still be associated with poor neurocognitive outcomes. The visualization suggest that there is a broad set of ICP doses that is not addressed in the guidelines although associated with worse long-term neurological outcomes. In spite of this epidemiological work [66–68, 72], the concept of ICP dose is not currently used in the clinical practice to evaluate the patients clinical status.

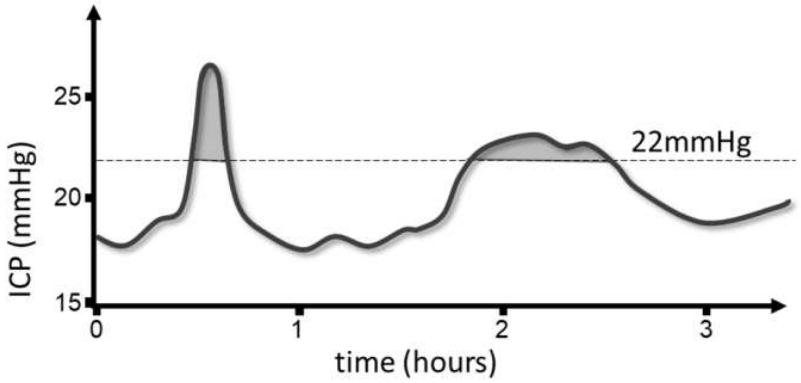


Figure 1.13 Visual representation of ICP dose. Example of dose of ICP above 22 mmHg, the dose of ICP, represented with the gray area, is the total area where ICP is above a certain threshold for a certain duration.

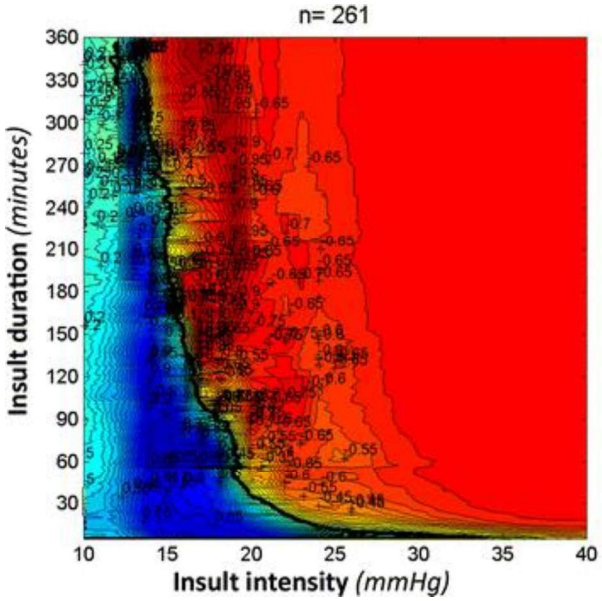


Figure 1.14 Visualization of the association between dose of ICP and long-term neurological outcomes in patients with TBI [66]. The plot shows, in a color-coded fashion, the association between events of at least a certain intensity X that lasts at least for a certain time Y and long-term neurological outcomes, quantified with the

Glasgow Outcome Score (GOS) at 6 months. Doses of ICP that occur more frequently in patients with better long-term neurological outcomes (higher GOS) are represented in blue, while points of ICP that occur more frequently in patients with worse long-term (lower GOS) outcomes are represented in red. A black exponential line, called transition curve, divides the two regions.

Given the complexity of TBI pathophysiology, ICP alone cannot provide complete knowledge of the status of the brain [62, 73]. For instance, the mean arterial blood pressure (MAP) is an important contributor to the perfusion status of the brain. Similarly, local brain oxygenation can be monitored with the partial brain tissue oxygen ($PbtO_2$), which monitors the balance between oxygen delivery and consumption. Moreover, cerebrovascular autoregulation (CAR) plays a crucial role in guaranteeing oxygen and nutrient supply. CAR is a dynamic phenomenon that is difficult to assess in a clinical context [74] and for which computational indices, such as the Pressure Reactivity index (PRx) [75] or the low-frequency autoregulatory index (LAx) [76], have been proposed. Both indexes are calculated as moving correlation between the MAP and ICP signals, a PRx above 0.3 and a LAx above 0 indicate impaired CAR. Currently, none of these indexes is available at the bedside unless in the form of research software (ICM+, Cambridge Enterprise Ltd., UK).

For patients with TBI, there is an urgent demand for a better definition of ICP-targeted therapy, as well as a need to bring knowledge derived from previous research, such as the concept of ICP dose or computational indexes of CAR, to the bedside.

1.5.3 *Sub-arachnoid hemorrhage*

Aneurysmal subarachnoid hemorrhage (aSAH) is a type of stroke most frequently caused by the rupture of an aneurysm of one of the cerebral arteries, which results in the outflow of blood in the subarachnoid space [77]. aSAH has high mortality [78], and approximately half of the survivors are at risk of long-term physical and neurocognitive impairment [79]. The main risk factors for poor outcomes are the patient's level of consciousness at ICU admission and the amount of sub-arachnoid blood as quantified by an initial computed tomography (CT) scan of the head [80]. The Modified Fisher Scale is used to score the sub-arachnoid and intraventricular blood [81]. Two main severity scales are used to grade level

of consciousness, the Hunt and Hess [82] and the World Federation of Neurosurgical Surgeons [83].

The ruptured aneurysm needs to be secured as soon as possible to prevent re-bleeding, which can occur in up to 7% of patients [84, 85]. Re-bleeding has a mortality rate of around 50% and is an important cause of poor long-term neurological outcomes [84]. Treatment options for aneurysms are neurosurgical clipping and endovascular coiling, depending on the age, clinical stability of the patient and the location of the aneurysm [77]. Once the aneurysm is secured, clinical attention is focused at preventing secondary complications [77, 85], such as vasospasm (46% of patients) [85, 86], hydrocephalus (20% of patients) [87] and increased ICP (81% of patients) [88, 89].

To prevent vasospasm, hypertension is allowed after securing the aneurysm, although the safe range remains unclear [77]. Hyperglycemia and hyperthermia should be corrected, given their association with poor outcomes [77, 90]. Prophylaxis of deep venous thrombosis should be started as early as possible [77, 90]. Calcium antagonists can be prescribed to reduce the risk of ischemic complications [77, 90]. External ventricular or lumbar drain can be used to treat hydrocephalus [77, 90].

Regarding the prevention and management of increased ICP, recommendations for patients with SAH [90] are based on the guidelines for patients with severe TBI, with the underlying hypothesis that the treatment threshold that is valid for patients with TBI can be equally applied to patients with SAH. As a result, similarly to TBI, also in patients with SAH aggressive treatment for elevated ICP is initiated when the ICP rises above 20-22 mmHg. However, there is no robust data to support this hypothesis. Although elevated ICP in patients with SAH is associated with increased mortality and poor long-term outcomes [88, 89, 91], safe ICP levels for patients with SAH remain unknown and specific recommendations for ICP management in SAH are missing. As a result of these uncertainties, treatment protocols for the management of elevated ICP in patients with SAH highly differ between centers. For example, ICP is only monitored in some centers, and only in the most severe cases of SAH. As previously explained, ICP monitoring is normally performed through an external ventricular drain and an intraparenchymal probe.

BIBLIOGRAPHY

1. Adhikari NK, Fowler RA, Bhagwanjee S, Rubenfeld GD (2010) Critical care and the global burden of critical illness in adults. Lancet 376:1339–1346. https://doi.org/10.1016/S0140-6736(10)60446-1

2. Joffe R, Duff J, Garcia Guerra G, et al (2016) The accuracy of blood pressure measured by arterial line and non-invasive cuff in critically ill children. Crit Care 20:177. https://doi.org/10.1186/s13054-016-1354-x

3. Ricci Z, Brogi J, De Filippis S, et al (2019) Arterial Pressure Monitoring in Pediatric Patients Undergoing Cardiac Surgery: An Observational Study Comparing Invasive and Non-invasive Measurements. Pediatr Cardiol 40:1231–1237. https://doi.org/10.1007/s00246-019-02137-9

4. Angus DC (2015) Fusing Randomized Trials With Big Data. JAMA 314:767. https://doi.org/10.1001/jama.2015.7762

5. Murdoch B T, Detsky S A (2013) The inevitable application of big data to health care. JAMA J Am Med Assoc 309:1351–1352. https://doi.org/10.1001/jama.2013.393

6. Obermeyer Z, Emanuel EJ (2016) Predicting the Future-Big Data, Machine Learning, and Clinical Medicine HHS Public Access. N Engl J Med 375:1216–1219. https://doi.org/10.1056/NEJMp1606181

7. Shickel B, Loftus TJ, Adhikari L, et al (2019) DeepSOFA: A Continuous Acuity Score for Critically Ill Patients using Clinically Interpretable Deep Learning. Sci Rep 9:1879. https://doi.org/10.1038/s41598-019-38491-0

8. Flechet M, Falini S, Bonetti C, et al (2019) Machine learning versus physicians' prediction of acute kidney injury in critically ill adults: a prospective evaluation of the AKIpredictor. Crit Care 23:282. https://doi.org/10.1186/s13054-019-2563-x

9. Frizzell JD, Liang L, Schulte PJ, et al (2017) Prediction of 30-Day All-Cause Readmissions in Patients Hospitalized for Heart Failure. JAMA Cardiol 2:204. https://doi.org/10.1001/jamacardio.2016.3956

10. Kansagara D, Englander H, Salanitro A, et al (2011) Risk Prediction Models for Hospital Readmission. JAMA 306:1688. https://doi.org/10.1001/jama.2011.1515

11. Rojas JC, Carey KA, Edelson DP, et al (2018) Predicting Intensive Care Unit Readmission with Machine Learning Using Electronic Health Record Data. Ann Am Thorac Soc 15:846–853. https://doi.org/10.1513/AnnalsATS.201710-787OC

12. Wong A, Young AT, Liang AS, et al (2018) Development and Validation of an Electronic Health Record–Based Machine Learning Model to Estimate Delirium Risk in Newly Hospitalized Patients Without Known Cognitive Impairment. JAMA Netw Open 1:e181018.

https://doi.org/10.1001/jamanetworkopen.2018.1018

13. Nemati S, Holder A, Razmi F, et al (2018) An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU. Crit Care Med 46:547–553. https://doi.org/10.1097/CCM.0000000000002936

14. Seymour CW, Kennedy JN, Wang S, et al (2019) Derivation, Validation, and Potential Treatment Implications of Novel Clinical Phenotypes for Sepsis. JAMA 321:2003. https://doi.org/10.1001/jama.2019.5791

15. Güiza F, Depreitere B, Piper I, et al (2013) Novel Methods to Predict Increased Intracranial Pressure During Intensive Care and Long-Term Neurologic Outcome After Traumatic Brain Injury: Development and Validation in a Multicenter Dataset. Crit Care Med 41:554–564. https://doi.org/DOI 10.1097/CCM.0b013e3182742d0a

16. Fleuren LM, Thoral P, Shillan D, et al (2020) Machine learning in intensive care medicine: ready for take-off? Intensive Care Med 46:1486–1488. https://doi.org/10.1007/s00134-020-06045-y

17. Vollmer S, Mateen BA, Bohner G, et al (2020) Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. BMJ l6927. https://doi.org/10.1136/bmj.l6927

18. Fenech M, Strukelij N, Buston O (2018) Ethical, Social, and Political challenges of Artificail Intelligence in health. https://wellcome.ac.uk/sites/default/files/ai-in-health-ethical-social-political-challenges.pdf

19. Cruz Rivera S, Liu X, Chan A-W, et al (2020) Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. Nat Med 26:1351–1363. https://doi.org/10.1038/s41591-020-1037-7

20. Liu X, Cruz Rivera S, Moher D, et al (2020) Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Nat Med 26:1364–1374. https://doi.org/10.1038/s41591-020-1034-x

21. Moons KGM, Altman DG, Reitsma JB, et al (2015) Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): Explanation and elaboration. Ann Intern Med 162:W1–W73. https://doi.org/10.7326/M14-0698

22. Doshi-Velez F, Kim B (2017) Towards A Rigorous Science of Interpretable Machine Learning

23. Steyerberg EW (2019) Clinical Prediction Models. Springer International Publishing, Cham

24. Tibshirani R (1996) Regression Shrinkage and Selection via the Lasso. J R

Stat Soc 267–288

25. Muthukrishnan R, Rohini R (2016) LASSO: A feature selection technique in predictive modeling for machine learning. In: 2016 IEEE International Conference on Advances in Computer Applications (ICACA). IEEE, pp 18–20

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

27. Shannon CE (1948) A Mathematical Theory of Communication. Bell Syst Tech J 27:379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

28. Rasmussen CE (2006) Gaussian Processes for Machine Learning

29. Loh W (2011) Classification and regression trees. WIREs Data Min Knowl Discov 1:14–23. https://doi.org/10.1002/widm.8

30. Leo Breiman (2001) Random Forests. Mach Learn 45:5–32

31. Sofaer HR, Hoeting JA, Jarnevich CS (2019) The area under the precision-recall curve as a performance metric for rare binary events. Methods Ecol Evol 10:565–577. https://doi.org/10.1111/2041-210X.13140

32. Austin PC, Steyerberg EW (2014) Graphical assessment of internal and external calibration of logistic regression models by using loess smoothers. Stat Med 33:517–535. https://doi.org/10.1002/sim.5941

33. Finazzi S, Poole D, Luciani D, et al (2011) Calibration Belt for Quality-of-Care Assessment Based on Dichotomous Outcomes. PLoS One 6:e16110. https://doi.org/10.1371/journal.pone.0016110

34. Vickers AJ, Elkin EB (2008) Decision curve analysis: a novel method for evaluating prediction models. Med Decis Mak 26:565–574. https://doi.org/10.1177/0272989X06295361.Decision

35. Hoffman JI., Kaplan S (2002) The incidence of congenital heart disease. J Am Coll Cardiol 39:1890–1900. https://doi.org/10.1016/S0735-1097(02)01886-7

36. Glass TJA, Seed M, Chau V (2019) Chapter 15 - Congenital Heart Disease: An Important Cause of Brain Injury and Dysmaturation, Third Edit. Elsevier Inc.

37. Oster ME, Lee KA, Honein MA, et al (2013) Temporal Trends in Survival Among Infants With Critical Congenital Heart Defects. Pediatrics 131:e1502–e1508. https://doi.org/10.1542/peds.2012-3435

38. Limperopoulos C, Tworetzky W, McElhinney DB, et al (2010) Brain Volume and Metabolism in Fetuses With Congenital Heart Disease. Circulation 121:26–33. https://doi.org/10.1161/CIRCULATIONAHA.109.865568

39. Schellen C, Ernst S, Gruber GM, et al (2015) Fetal MRI detects early alterations of brain development in Tetralogy of Fallot. Am J Obstet Gynecol 213:392.e1-392.e7. https://doi.org/10.1016/j.ajog.2015.05.046

40. Sun L, Macgowan CK, Sled JG, et al (2015) Reduced Fetal Cerebral Oxygen Consumption Is Associated With Smaller Brain Size in Fetuses With Congenital Heart Disease. Circulation 131:1313–1323. https://doi.org/10.1161/CIRCULATIONAHA.114.013051

41. McQuillen PS, Barkovich AJ, Hamrick SEG, et al (2007) Temporal and Anatomic Risk Profile of Brain Injury With Neonatal Repair of Congenital Heart Defects. Stroke 38:736–741. https://doi.org/10.1161/01.STR.0000247941.41234.90

42. Andropoulos DB, Hunter J V., Nelson DP, et al (2010) Brain immaturity is associated with brain injury before and after neonatal cardiac surgery with high-flow bypass and cerebral oxygenation monitoring. J Thorac Cardiovasc Surg 139:543–556. https://doi.org/10.1016/j.jtcvs.2009.08.022

43. Galli KK, Zimmerman RA, Jarvik GP, et al (2004) Periventricular leukomalacia is common after neonatal cardiac surgery. J Thorac Cardiovasc Surg 127:692–704. https://doi.org/10.1016/j.jtcvs.2003.09.053

44. Hayashida M, Kin N, Tomioka T, et al (2004) Cerebral ischaemia during cardiac surgery in children detected by combined monitoring of BIS and near-infrared spectroscopy. Br J Anaesth 92:662–669. https://doi.org/10.1093/bja/aeh120

45. Spaeder MC, Klugman D, Skurow-Todd K, et al (2017) Perioperative Near-Infrared Spectroscopy Monitoring in Neonates with Congenital Heart Disease: Relationship of Cerebral Tissue Oxygenation Index Variability with Neurodevelopmental Outcome. Pediatr Crit Care Med 18:213–218. https://doi.org/10.1097/PCC.0000000000001056

46. Desmond F, Namachivayam S (2016) Does near-infrared spectroscopy play a role in paediatric intensive care? BJA Educ 16:281–285. https://doi.org/10.1093/bjaed/mkv053

47. Smith-Wright DL, Green TP, Lock JE, et al (1984) Complications of vascular catheterization in critically ill children. Crit Care Med 12:1015–1017. https://doi.org/10.1097/00003246-198412000-00001

48. Petäjä J, Lundström U, Sairanen H, et al (1996) Central venous thrombosis after cardiac operations in children. J Thorac Cardiovasc Surg 112:883–889. https://doi.org/10.1016/S0022-5223(96)70087-9

49. Ghanayem NS, Wernovsky G, Hoffman GM (2011) Near-infrared spectroscopy as a hemodynamic monitor in critical illness. Pediatr Crit Care Med 12:S27–S32. https://doi.org/10.1097/PCC.0b013e318221173a

50. Zulueta JL, Vida VL, Perisinotto E, et al (2013) The Role of Intraoperative Regional Oxygen Saturation Using Near Infrared Spectroscopy in the Prediction of Low Output Syndrome After Pediatric Heart Surgery. J Card Surg 28:446–452. https://doi.org/10.1111/jocs.12122

51. Pellicer A, Bravo M del C (2011) Near-infrared spectroscopy: A methodology-focused review. Semin Fetal Neonatal Med 16:42–49. https://doi.org/10.1016/j.siny.2010.05.003

52. Neshat Vahid S, Panisello JM (2014) The state of affairs of neurologic monitoring by near-infrared spectroscopy in pediatric cardiac critical care. Curr Opin Pediatr 26:299–303. https://doi.org/10.1097/MOP.0000000000000098

53. Scheeren TWL, Schober P, Schwarte LA (2012) Monitoring tissue oxygenation by near infrared spectroscopy (NIRS): background and current applications. J Clin Monit Comput 26:279–287. https://doi.org/10.1007/s10877-012-9348-y

54. Koch JD, Kernie SG (2011) Protecting the future: neuroprotective strategies in the pediatric intensive care unit. Curr Opin Pediatr 23:275–280. https://doi.org/10.1097/MOP.0b013e3283460584

55. Hoffman GM, Brosig CL, Mussatto KA, et al (2013) Perioperative cerebral oxygen saturation in neonates with hypoplastic left heart syndrome and childhood neurodevelopmental outcome. J Thorac Cardiovasc Surg 146:1153–1164. https://doi.org/10.1016/j.jtcvs.2012.12.060

56. Sood ED, Benzaquen JS, Davies RR, et al (2013) Predictive value of perioperative near-infrared spectroscopy for neurodevelopmental outcomes after cardiac surgery in infancy. J Thorac Cardiovasc Surg 145:. https://doi.org/10.1016/j.jtcvs.2012.10.033

57. Maas AIR, Menon DK, Adelson PD, et al (2017) Traumatic brain injury: integrated approaches to improve prevention, clinical care, and research. Lancet Neurol 4422:. https://doi.org/10.1016/S1474-4422(17)30371-X

58. Steyerberg EW, Wiegers E, Sewalt C, et al (2019) Case-mix, care pathways, and outcomes in patients with traumatic brain injury in CENTER-TBI: a European prospective, multicentre, longitudinal, cohort study. Lancet Neurol 18:923–934. https://doi.org/10.1016/S1474-4422(19)30232-7

59. Stocchetti N, Maas AIR (2014) Traumatic Intracranial Hypertension. N Engl J Med 370:2121–2130. https://doi.org/10.1056/NEJMra1208708

60. Hawryluk GWJ, Aguilera S, Buki A, et al (2019) A management algorithm for patients with intracranial pressure monitoring: the Seattle International Severe Traumatic Brain Injury Consensus Conference (SIBICC). Intensive Care Med 45:1783–1794. https://doi.org/10.1007/s00134-019-05805-9

61. Marmarou A, Anderson RL, Ward JD, et al (1991) Impact of ICP instability

and hypotension on outcome in patients with severe head trauma. J Neurosurg 75:S59–S66. https://doi.org/10.3171/sup.1991.75.1s.0s59

62. Olson DM, Andrew Kofke W, O'Phelan K, et al (2015) Global Monitoring in the Neurocritical Care Unit. Neurocrit Care 22:337–347. https://doi.org/10.1007/s12028-015-0132-y

63. Carney N, Totten AM, O'Reilly C, et al (2016) Guidelines for the Management of Severe Traumatic Brain Injury 4th Edition

64. Helbok R, Meyfroidt G, Beer R (2018) Intracranial pressure thresholds in severe traumatic brain injury: Con: The injured brain is not aware of ICP thresholds! Intensive Care Med 44:1318–1320. https://doi.org/10.1007/s00134-018-5249-y

65. Stocchetti N, Poole D, Okonkwo DO (2018) Intracranial pressure thresholds in severe traumatic brain injury: we are not sure: Prudent clinical practice despite dogma or nihilism. Intensive Care Med 44:1321–1323. https://doi.org/10.1007/s00134-018-5251-4

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

67. Vik A, Nag T, Fredriksli O, et al (2008) Relationship of " dose " of intracranial hypertension to outcome in severe traumatic brain injury. J Neurosurg J Neurosurg J Neurosurg 109:0–678. https://doi.org/10.3171/JNS/2008/109/10/0678

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

69. Sheth KN, Stein DM, Aarabi B, et al (2013) Intracranial Pressure Dose and Outcome in Traumatic Brain Injury. Neurocrit Care 18:26–32. https://doi.org/10.1007/s12028-012-9780-3

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

71. Lazaridis C, DeSantis SM, Smielewski P, et al (2014) Patient-specific thresholds of intracranial pressure in severe traumatic brain injury. J Neurosurg 120:893–900. https://doi.org/10.3171/2014.1.JNS131292

72. Donnelly J, Güiza F, Depreitere B, et al (2021) Visualising the pressure-time burden of elevated intracranial pressure after severe traumatic brain injury: a retrospective confirmatory study. Br J Anaesth 126:e15–e17. https://doi.org/10.1016/j.bja.2020.09.018

73. Citerio G, Oddo M, Taccone FS (2015) Recommendations for the use of multimodal monitoring in the neurointensive care unit. Curr Opin Crit Care 21:113–119. https://doi.org/10.1097/MCC.0000000000000179

74. Depreitere B, Citerio G, Smith M, et al (2021) Cerebrovascular Autoregulation Monitoring in the Management of Adult Severe Traumatic Brain Injury: A Delphi Consensus of Clinicians. Neurocrit Care 34:731–738. https://doi.org/10.1007/s12028-020-01185-x

75. Czosnyka M, Smielewski P, Kirkpatrick P, et al (1997) Continuous assessment of the cerebral vasomotor reactivity in head injury. Neurosurgery 41:11–7; discussion 17-9

76. Depreitere B, Güiza F, Van den Berghe G, et al (2014) Pressure autoregulation monitoring and cerebral perfusion pressure target recommendation in patients with severe traumatic brain injury based on minute-by-minute monitoring data. J Neurosurg 120:1451–1457. https://doi.org/10.3171/2014.3.JNS131500

77. Suarez J, Tarr R, Selman W (2006) Aneurysmal subarachnoid hemorrhage. N Engl J Med 354:387–396. https://doi.org/10.1056/NEJMra052732

78. van Gijn J, Rinkel GJE (2001) Subarachnoid haemorrhage: diagnosis, causes and management. Brain 124:249–278. https://doi.org/10.1093/brain/124.2.249

79. Hackett ML, Anderson CS (2000) Health outcomes 1 year after subarachnoid hemorrhage: An international population-based study. Neurology 55:658–662. https://doi.org/10.1212/WNL.55.5.658

80. Hijdra A, van Gijn J, Nagelkerke NJ, et al (1988) Prediction of delayed cerebral ischemia, rebleeding, and outcome after aneurysmal subarachnoid hemorrhage. Stroke 19:1250–1256. https://doi.org/10.1161/01.STR.19.10.1250

81. Frontera JA, Claassen J, Schmidt JM, et al (2006) Prediction of Symptomatic Vasospasmafter Subarachnoid Hemorrhage: The Modified Fisher Scale. Neurosurgery 59:21–27. https://doi.org/10.1227/01.NEU.0000218821.34014.1B

82. Hunt WE, Hess RM (1968) Surgical Risk as Related to Time of Intervention in the Repair of Intracranial Aneurysms. J Neurosurg 28:14–20. https://doi.org/10.3171/jns.1968.28.1.0014

83. (1988) Report of World Federation of Neurological Surgeons Committee on a Universal Subarachnoid Hemorrhage Grading Scale. J Neurosurg 68:. https://doi.org/10.3171/jns.1988.68.6.0985

84. Naidech AM, Janjua N, Kreiter KT, et al (2005) Predictors and Impact of Aneurysm Rebleeding After Subarachnoid Hemorrhage. Arch Neurol 62:410. https://doi.org/10.1001/archneur.62.3.410

85. Solenski NJ, Haley ECJ, Kassell NF, et al (1995) Medical complications of aneurysmal subarachnoid hemorrhage. Crit Care Med 23:1007–1017. https://doi.org/10.1097/00003246-199506000-00004

86. Frontera JA, Fernandez A, Schmidt JM, et al (2009) Defining Vasospasm After Subarachnoid Hemorrhage. Stroke 40:1963–1968. https://doi.org/10.1161/STROKEAHA.108.544700

87. Germanwala A V., Huang J, Tamargo RJ (2010) Hydrocephalus After Aneurysmal Subarachnoid Hemorrhage. Neurosurg Clin N Am 21:263–270. https://doi.org/10.1016/j.nec.2009.10.013

88. Zoerle T, Lombardo A, Colombo A, et al (2015) Intracranial pressure after subarachnoid hemorrhage. Crit Care Med 43:168–176. https://doi.org/10.1097/CCM.0000000000000670

89. Magni F, Pozzi M, Rota M, et al (2015) High-Resolution Intracranial Pressure Burden and Outcome in Subarachnoid Hemorrhage. Stroke 46:2464–2469. https://doi.org/10.1161/STROKEAHA.115.010219

90. Connolly ES, Rabinstein AA, Carhuapoma JR, et al (2012) Guidelines for the Management of Aneurysmal Subarachnoid Hemorrhage. Stroke 43:1711–1737. https://doi.org/10.1161/STR.0b013e3182587839

91. Cossu G, Messerer M, Stocchetti N, et al (2016) Intracranial pressure and outcome in critically ill patients with aneurysmal subarachnoid hemorrhage: A systematic review. Minerva Anestesiol 82:684–696

2

OBJECTIVES

2.1 GENERAL AIM

The general aim of this thesis is to use artificial intelligence to gain insights on the complex association between neuro-monitoring signals and long-term neurocognitive outcomes in critically ill patients at risk of brain injuries and to develop accurate decision support applications for the use at the bedside. The ultimate goal of this thesis is to create knowledge that contributes to the line of research aimed at reducing the long-term neurocognitive burden of critical illness.

2.2 RESEARCH OBJECTIVES

Objective 1 The first objective of this thesis is to investigate the independent association between brain tissue oxygen saturation ($SctO_2$) and long-term neurocognitive outcomes in pediatric patients after cardiac surgery for congenital heart diseases. Using data from a prospective observational study, I will investigate whether post-operative reduced $SctO_2$ values and increased $SctO_2$ desaturation is associated with reduced total IQ at 2 years follow-up **(Chapter 3)**.

Objective 2 The second objective of this thesis is to investigate the independent association between elevated doses of intracranial pressure and long-term neurological outcomes in patients with sub-arachnoid hemorrhage. The independent association between ICP dose and outcomes will be investigated in a large multi-center cohort. Moreover, I will visually explore this association by using advanced visualization techniques from previous epidemiological studies **(Chapter 4)**.

Objective 3 The third objective of this thesis is the temporal and geographical validation of an existing machine learning model for the prediction of episodes of extremely elevated intracranial pressure in

patients with traumatic brain injury on a large, prospectively collected European dataset **(Chapter 5)**.

Objective 4 The fourth objective of this thesis is to develop a machine learning prediction model for the early detection of future episodes of harmful doses of intracranial pressure in patients with traumatic brain injury. To achieve this objective I will use two of the biggest European multi-center datasets of demographics and high-quality monitoring data of patients with traumatic brain injury **(Chapter 6)**.

Objective 5 The fifth objective of this thesis is to design, implement and validate through a prospective study, a bedside decision-support software prototype for the management of patients with traumatic brain injury. The software not only provides predictions of potentially harmful ICP events, but it also computes and displays relevant metrics for the management of patients with traumatic brain injury **(Chapter 7,8)**.

3

POST-OPERATIVE CEREBRAL OXYGEN SATURATION IN CHILDREN AFTER CONGENITAL CARDIAC SURGERY AND LONG-TERM TOTAL IQ: A PROSPECTIVE OBSERVATIONAL STUDY

ADAPTED FROM: **Carra G.** et al. "Post-operative cerebral oxygen saturation in children after congenital cardiac surgery and long-term total IQ: a prospective observational study". Critical Care Medicine, 2021 Jun 1; 49(6):967-976.

Presented as:

- Oral presentation at the 18th ICP2019 conference. Leuven, Belgium, September 2019.
- Oral presentation at the Belgian Society of Pediatric Neurology Autumn Meeting. Leuven, Belgium, October, 2019.
- Poster presentation at the 33rd European Society of Intensive Care Medicine (ESICM) Annual Congress. Online, December, 2020.
- Oral presentation at the Critical Care Webcast. Online, June, 2021.

ABSTRACT

PURPOSE: During the early post-operative period, children with congenital heart disease can suffer from inadequate cerebral perfusion, with possible long-term neurocognitive consequences. Cerebral tissue oxygen saturation ($SctO_2$) can be monitored non-invasively with near-infrared spectroscopy (NIRS). In this prospective study, we hypothesized that reduced $SctO_2$ and increased intensity and duration of desaturation (defined as $SctO_2$ < 65%) during the early post-operative period, independently increase the probability of reduced total IQ, 2-years after admission to a Pediatric Intensive Care Unit (PICU).

METHODS: The study included pediatric patients after surgery for congenital heart disease admitted to the PICU of the University Hospitals Leuven, Belgium between 2012 and 2015. Post-operative cerebral perfusion was characterized with the mean $SctO_2$ and dose of desaturation of the first 12 hours and 24 hours of $SctO_2$ monitoring. The independent association of post-operative mean $SctO_2$ and dose of desaturation with total IQ at 2-years follow-up was evaluated with a Bayesian linear regression model adjusted for known confounders.

RESULTS: According to a non-informative prior, reduced mean $SctO_2$ during the first 12 hours of monitoring results in a loss of IQ points at 2-years, with a 90% probability (posterior β estimates [80% credible interval] 0.23 [0.04 to 0.41]). Similarly, increased dose of $SctO_2$ desaturation would result in a loss of IQ points at 2-years with a 90% probability (posterior β estimates [80% credible interval] -0.009 [-0.016 to -0.001]).

CONCLUSIONS: Increased dose of $SctO_2$ desaturation and reduced mean $SctO_2$ during the early post-operative period, independently increase the probability of having a lower total IQ, 2 years after PICU admission.

3.1 INTRODUCTION

Half of all children with congenital heart disease (CHD) undergo surgery in the first months of life [1, 2]. Despite high survival rates, many of the surviving children can suffer from long-term neurodevelopmental deficits [3–6], as a result of white matter lesions following hypoxic or ischemic events [2, 3, 7, 8]. To maximize outcomes, the treating team pursue adequate oxygen delivery during the entire perioperative period. Cerebral perfusion can be monitored with near-infrared spectroscopy (NIRS), which continuously monitors regional cerebral tissue oxygen saturation ($SctO_2$) in the frontal lobe [9, 10].

Perioperative reduced $SctO_2$ is associated with longer length of stay (LOS) in the pediatric intensive care unit (PICU), longer duration of mechanical ventilation and increased risk of major post-operative complications and mortality [11–16]. It remains unclear whether reduced $SctO_2$ is also associated with long-term neurocognitive deficits [14]. Several studies have been performed to address this question, but with contradictory results, due to the limited number of the patients enrolled and the lack of a uniform protocol to treat non-optimal $SctO_2$ [14].

In this study, we used a large single-center dataset to investigate the association between $SctO_2$ in pediatric patients after surgery for CHD and total IQ at 2-years follow-up. The associations were analyzed with Bayesian inference, which allows accounting for uncertainty in parameter estimates.

3.2 MATERIAL AND METHODS

3.2.1 *Study design*

This study includes children with CHD admitted in the PICU of the University Hospitals Leuven, Belgium, between 2012 and 2015. It derives from a protocol amendment to a blinded prospective observational study to investigate the independent association of continuous $SctO_2$ measurements in children admitted after congenital cardiac surgery, on organ failure, PICU and hospital outcomes [13, 17]. Seventy-nine (91%) of the children included in this study were also included in the PEPaNIC trial [18].

The protocol of the study was pre-registered (ClinicalTrials.gov: NCT01706497). Both the original study and the amendment obtained ethical approval of the ethics committee research UZ / KU Leuven (EC research).

3.2.2 *Participants*

Pediatric patients younger than 12 years of age, admitted to the PICU of the University Hospitals Leuven, Belgium, who underwent corrective or palliative cardiac surgery for a congenital heart defect, were eligible for the study. A complete list of the cardiac surgery procedures amenable for inclusion is reported in Appendix 3.A.1. The enrolled children had to be mechanically ventilated upon PICU admission, have an arterial catheter in place, and have an expected PICU length of stay longer than 24 hours. Patients with actual or suspected brain damage, e.g. children who had cardiopulmonary resuscitation, severe traumatic brain injury or chronic neuropathy, were excluded from the study. Additional reasons for exclusion were the presence of clinical or physiologic conditions that prevented the correct placement of the NIRS sensors on the patients' forehead or further admissions to PICU before the long-term follow-up. Baseline characteristics including age, gender, educational level of the parents (as described in Appendix 3.A.2), severity of illness (Pediatric Index of Mortality 3 (PIM3) score), presence of cyanosis before and after surgery, and predefined syndrome (as described in Appendix 3.A.3) were recorded upon PICU admission. The nutrition strategy (early or late initiation of parenteral nutrition, as described in Appendix 3.A.4) was recorded. Surgery information such as the Risk Adjustment for Congenital Heart Surgery 1 (RACHS-1) score, whether the child underwent cardiopulmonary bypass (CPB) or deep hypothermic circulatory arrest (DHCA) and duration of CPB were recoded.

3.2.3 *Cerebral NIRS Monitoring*

$SctO_2$ was monitored continuously using the FORESIGHT cerebral oximeter (CAS Medical Systems, Branford, CT. The unit of measurement of $SctO_2$ is in %). The attending bedside clinicians did not have access to the NIRS data in order not to influence the independent predictive value of the signal. For that purpose, the monitor screens were blinded with a sealed screen cover. Monitoring data were stored in the Patient Data Management System (MetaVision; iMD-Soft, Needham, MA), with a minute-by-minute time resolution. The signals were assessed through visual inspection by an experienced clinicians and considered of "low-quality" if, among others, the signal was absent or presented clear artifacts for more than 50% of the monitoring time or if it was shorter than 2 hours.

3.2.4 *Neurocognitive outcomes*

Total IQ of the included children was assessed 2-years after PICU admission by trained psychologists with the Wechsler Intelligence Quotient (IQ) tests [19, 20]. Namely, the WPPSI-III-NL [19] was used for children aged between 2.5 years and 5 years and 11 months, while the Wechsler Intelligence Scale for Children (WISC-III-NL) [20] was used for children aged between 6 years and 16 years and 11 months. Seventy-nine of the included children were assessed in the framework of the PEPaNIC study [18], see Appendix 3.A.5-3.A.6. The children were assessed at the hospital or at home. The second option was offered to the parents or caregivers not able or not willing to travel to the hospital.

3.2.5 *$SctO_2$ predictors*

The $SctO_2$ traces obtained from the left and right electrodes were averaged. Obvious artifacts ($SctO_2$ values below 20%) were removed and single missing values were imputed through linear interpolation. Desaturation was defined as $SctO_2$ below 65%. Mean $SctO_2$ and dose of desaturation were computed from the continuous $SctO_2$ signal as possible predictors of lower total IQ at 2-years follow-up [13, 15, 16]. The dose of desaturation combines the intensity and the duration of $SctO_2 < 65\%$ (unit of measurement in %·minutes), a detailed description of the computation method is included in Appendix 3.A.7. An example of dose of desaturation is visualized in Figure 3.1 with the dark-grey area between the signal and the grey dashed line, which indicates the 65% desaturation threshold. The $SctO_2$ predictors were extracted from the beginning of the monitoring up to the first 12 hours and up to the first 24 hours of $SctO_2$ monitoring. To be included, children had to have at least 2 hours of continuous $SctO_2$ monitoring, to account for the variation in duration of the $SctO_2$ recordings the dose was computed as percentage.

Children who experienced prolonged hyperoxia ($SctO_2 > 85\%$ for more than 10% of the entire monitoring time) were excluded from the analysis.

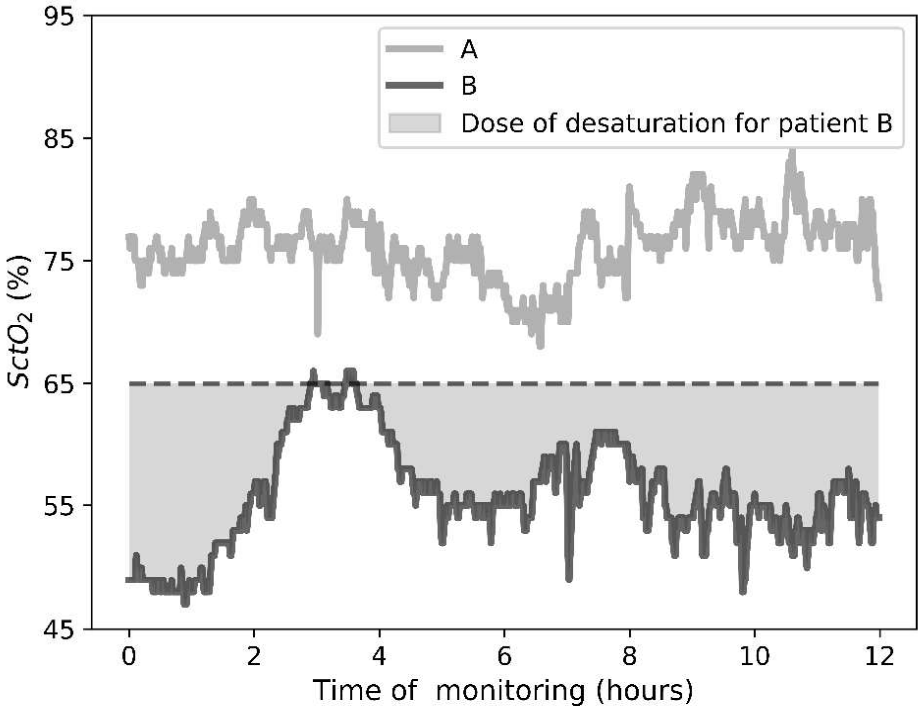


Figure 3.1 Sequences of SctO2 recordings, first 12 hours of monitoring time. Signal A desaturation dose: 0%·minutes, mean SctO2: 74%. Signal B desaturation dose: 917 %·minutes; mean SctO2: 57%. For signal B, the desaturation dose below 65% (unit of measurement: %·minutes) is represented by the dark-grey area between the signal and the grey dashed line which indicates the 65% desaturation threshold. According to the enthusiastic prior, there is an 80% probability that the dose of desaturation of signal B results in a loss of 4 to 15 IQ points as compared with signal A. Similarly, there is an 80% probability that the mean SctO2 of signal B results in 2.8 to 7.4 IQ points lower total IQ than signal A 2-years after pediatric intensive care medicine admission.

3.2.6 *Statistical analysis*

The independent association between $SctO_2$ predictors and total IQ was investigated through Bayesian multivariable linear regression models. In Bayesian analysis, the final β estimates of the linear regression model (posterior β estimates), integrate prior knowledge (results of previous studies or common knowledge) with the information provided by the analyzed data.

Three models with different types of prior probability distributions for the $SctO_2$ predictors under analysis were built: a) neutral prior distribution: it is a non-informative prior, assumes no association between the $SctO_2$

predictors and outcomes; b) enthusiastic prior distribution: it assumes that desaturation is associated with lower total IQ; c) skeptical prior distribution: it assumes that desaturation is associated with higher total IQ. The standard deviations of the skeptical and enthusiastic priors, which quantify the uncertainty of the parameter, were set to respectively twice and half their prior most likely value.

In Bayesian statistics, the credible interval of the posterior distribution indicates the range of values to which the posterior β estimate belongs with a certain probability. In this study, the 80% credible interval was used.

The model was adjusted for co-factors that have been shown to affect neurocognitive outcomes in previous studies, namely: age, nutrition strategy, presence of syndrome, presence of cyanosis after surgery and PIM3 score [18, 21–27]. The model was adjusted for nutrition strategy since part of the included children were also included in the PEPaNIC trial [18], which aimed at investigating the effect of early versus late administration of parenteral nutrition on the short and long-term outcomes of the children involved. Detailed information on the principles and construction of the Bayesian model, including informative priors set for the presence of syndrome and presence of cyanosis, are reported in Appendix 3.A.8.

Results of the Bayesian model were compared to a frequentist multivariable linear regression model (traditional statistical method), adjusted for the same co-factors of the Bayesian model. Statistical significance was set at a *p-value* of 0.05.

A sensitivity analysis for the definition of desaturation was performed, using as desaturation thresholds 60% and 55%. Moreover, two sensitivity analyses were performed on the effect of additional adjusting factors. The multivariable linear model was additionally adjusted for the educational level of the parents and, in a separate analysis, for the duration of CPB. The analyses were limited to the subsets of patients that had the required information. In addition, an interaction effect analysis was performed to investigate the effect of presence of cyanosis after surgery on $SctO_2$ mean and dose of desaturation.

Ordinal and continuous variables were reported using mean and SD for normal distributions and using median and 25^{th}-75^{th} interquartile (IQR) range for non-normal distributions.

Analyses were performed using Python version 2.7 (Python Software Foundation, http://www.python.org), Scipy version 0.19 (https://www.scipy.org), and Pymc3 version 3.5 (https://docs.pymc.io).

3.3 RESULTS

3.3.1 *Study population*

Three hundred children were recruited for the study. Of the enrolled patients, 78 patients were lost at follow-up, of the remaining 222 patients, 126 patients had low-quality NIRS and 9 died. Eighty-seven patients were included in the analysis, see flow diagram for patient inclusion in Figure 3.2.

Table 3.1 reports baseline demographics and clinical characteristics in PICU of the children included in the study. At PICU admission, the median (IQR) age of the children was 4 (1 to 13) months. The median (IQR) PIM3 was -3.37 (-3.93 to -2.67). A syndromic diagnosis was present in 8 (9%) of patients, and 27 (31%) patients had presence of cyanosis after surgery. Parenteral nutrition was withheld in the first week of PICU stay in 43 (49%) patients. Children had a mean (SD) 2-years follow-up total IQ of 91.9 (13.7). The mean $SctO_2$ for both the first 12 hours and 24 hours of monitoring time was equal to a mean (SD) of 71 (7) %, while the dose of desaturation was equal to a median (IQR) of 0.42 (0 to 53) %·minutes and 0.64 (0 to 49) %·minutes for the first 12 hours and 24 hours of monitoring time, respectively.

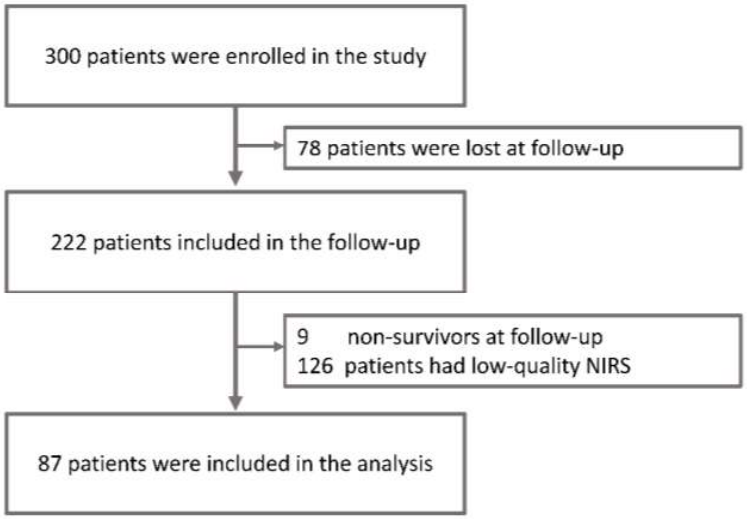


Figure 3.2 Flow diagram for study participants inclusion.

Table 3.1 Baseline demographics, clinical characteristics and outcomes of participating children

CHARACTERISTICS	DESCRIPTIVE STATISTICS N=87
DEMOGRAPHICS	
Age (months), median (IQR)	4 (1 to 13)
Gender (male), n (%)	56 (64)
EDUCATIONAL LEVEL OF THE PARENTS [a]	
Level 1, n (%)	3 (3)
Level 1.5, n (%)	7 (8)
Level 2, n (%)	19 (22)
Level 2.5, n (%)	20 (23)
Level 3, n (%)	27 (31)
Unknown, n (%)	11 (13)
CLINICAL SITUATION AT ADMISSION	
PIM3 score, median (IQR)	-3.34 (-3.93 to -2.67)
PIM3 probability of death (%), median (IQR)	3 (2 to 6)
Patients with syndrome, n (%)	8 (9)
Children with cyanosis before surgery, n (%)	54 (62)
SURGERY	
RACHS-1 score, median (IQR)	2 (2 to 3)
Patients that underwent CPB, n (%)	78 (89)
Patients that underwent DHCA, n (%)	3 (4)
CPB duration [b] (minutes), median (IQR)	77 (58 to 109)
POSTOPERATIVE CLINICAL SITUATION	
Children with persistent cyanosis after surgery, n (%)	27 (31)
Nutrition strategy (late parenteral nutrition), n (%)	43 (49)
Duration of $SctO_2$ monitoring (hours), median (IQR)	16 (9 to 28)
PICU LENGTH OF STAY (DAYS), MEAN (SD)	7 (10)
TOTAL IQ, MEAN (IQR)	91.9 (13.7)
$SctO_2$ FEATURES, FIRST 12 HOURS OF MONITORING	
$SctO_2$ mean (%), mean (SD)	71 (7)

$SctO_2$ dose of desaturation (%·minutes), median (IQR)	0.42 (0 to 53)
$SctO_2$ FEATURES, FIRST 24 HOURS OF MONITORING	
$SctO_2$ mean (%), mean (SD)	71 (7)
$SctO_2$ dose of desaturation (%·minutes), median (IQR)	0.64 (0 to 49)

Legend Table 3.1: PIM3 = Pediatric Index of Mortality 3. RACHS-1 = Risk Adjustment for Congenital Heart Surgery 1. CPB = cardiopulmonary bypass. DHCA = deep hypothermic circulatory arrest. PICU= Pediatric Intensive Care Unit.

[a] the educational level of the parents was computed as the mean of the level of the maternal and paternal educational levels. The single score is based on a 3-point scale (1 is low, 2 is middle, and 3 is high; see Appendix 3.A.2).

[b] data available only for children that underwent CPB, the statistic is computed on this subset of children

3.3.2 *Association between SctO2 predictors and 2-year neurocognitive development*

Bayesian analysis using neutral, skeptical and enthusiastic priors provided consistent results. In fact in each analysis an increased dose of $SctO_2$ desaturation and a reduced mean $SctO_2$ during the first 12 and 24 hours of $SctO_2$ monitoring, independently increased the probability of lower total IQ at 2-years follow-up, results are reported in Table 3.2 as posterior β estimates [80% credible interval]. The credible interval indicates the range of β estimates at which the posterior β value belongs with an 80% probability. In addition, Table 3.2 reports the probability that the posterior β estimate is strictly negative (or strictly positive). An increased dose of $SctO_2$ desaturation during the first 12 hours of monitoring resulted in a loss of IQ points at 2-years follow-up with a 90%, 90% and 97.5% probability, respectively, for the skeptical, neutral and enthusiastic prior (posterior β estimates [80% credible interval] equal to -0.007 [-0.015 to 0.000], -0.009 [-0.016 to -0.001] and -0.011 [-0.017 to -0.005], respectively). The corresponding results from the frequentist approach (traditional statistical method) did not show a statistically significant association (β estimates [CI]

per %·minutes increase equal to -0.009 [-0.021 to 0.003], R^2: 0.19, *p-value*: 0.14). Reduced mean $SctO_2$ during the first 12 hours of monitoring resulted in lower total IQ at 2-years follow-up with a 90%, 90% and 97.5% probability for the skeptical, neutral and enthusiastic priors, respectively (posterior β estimates [80% credible interval] equal to 0.208 [0.015 to 0.412] , 0.227 [0.037 to 0.412] and 0.300 [0.166 to 0.436], respectively). A representation of the priors and posterior probability distributions for the mean $SctO_2$ is shown in Figure 3.3. The corresponding results from the frequentist approach did not show a statistically significant association (β estimates [CI] per % increase equal to 0.36 [-0.084 to 0.811], R^2: 0.20, *p-value*: 0.11). Similar results were obtained when taking into account the first 24 hours of monitoring (Table 3.2). Figure 3.1 shows an example of $SctO_2$ episodes associated with reduced neurocognitive outcomes. According to the enthusiastic prior, there is an 80% probability that the dose of desaturation of signal B (Figure 3.1) results in a loss of 4 to 15 IQ points as compared with signal A (Figure 3.1). Similarly, there is an 80% probability that the mean $SctO_2$ of signal B (Figure 3.1) results in 2.8 to 7.4 IQ points lower total IQ than signal A (Figure 3.1) 2-years after PICU admission. Additional examples of $SctO_2$ episodes associated with reduced neurocognitive outcomes are reported in Appendix 3.A.9. Posterior β estimates and representation of prior and posterior distributions for all co-variates are reported in Appendix 3.A.10.

Sensitivity analyses for the definition of desaturation provided consistent results with the previous analyses, see Appendix 3.A.11. Similarly, sensitivity analyses on the effect of additional adjusting factors to the multivariable linear model provided consistent results, see Appendix 3.A.12. No interaction effect was observed between presence of cyanosis, mean $SctO_2$ and desaturation, see Appendix 3.A.13.

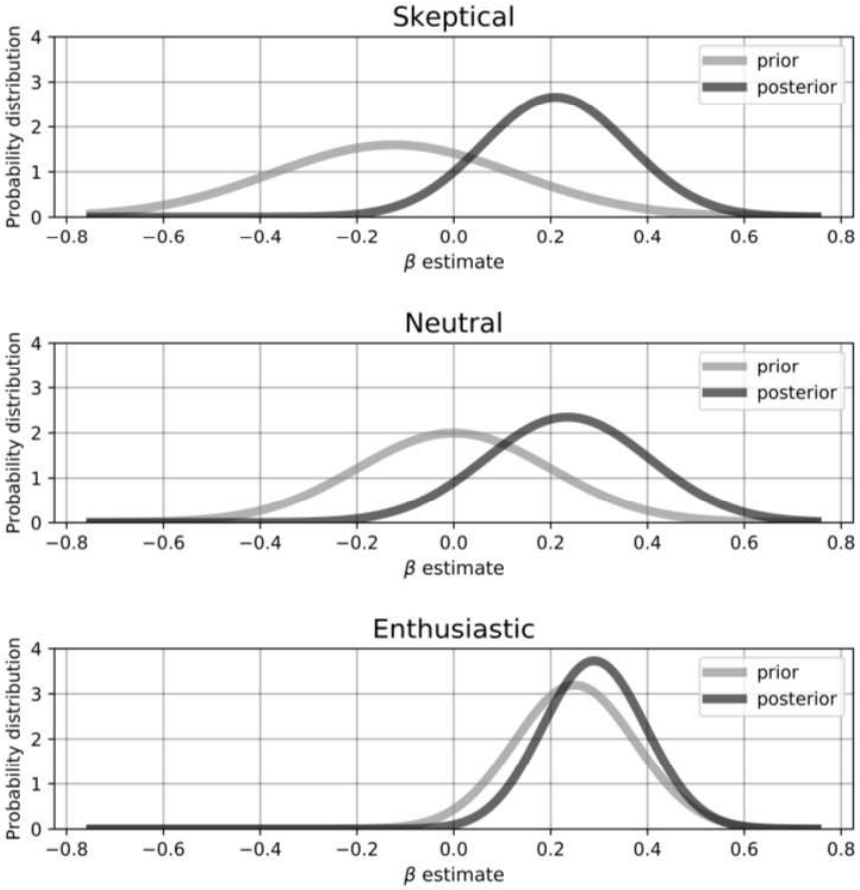


Figure 3.3 Comparison between the β priors and β posteriors probability distribution of the skeptical, neutral and enthusiastic priors of the multivariable Bayesian model on the relation between mean $SctO_2$ and total IQ. Only the distributions of the β estimate of the mean $SctO_2$ (first 12 hours of monitoring time) are shown.

Table 3.2 Multivariable linear regression analysis for the association between $SctO_2$ features and total IQ

VARIABLES	DESATURATION DOSE		MEAN $SctO_2$	
	Posterior β estimates [80% credible interval]	Probability that the posterior β estimate is strictly negative(-)/positive(+)	Posterior β estimates [80% credible interval]	Probability that the posterior β estimate is strictly negative(-)/positive(+)
First 12-hours of $SctO_2$ monitoring				
Skeptical prior	-0.007 [-0.015 to 0.000]	90% (-)	0.208 [0.015 to 0.412]	90% (+)
Neutral prior	-0.009 [-0.016 to -0.001]	90% (-)	0.227 [0.037 to 0.412]	90% (+)
Enthusiastic prior	-0.011 [-0.017 to -0.005]	97.5% (-)	0.300 [0.166 to 0.436]	97.5% (+)
First 24-hours of $SctO_2$ monitoring				
Skeptical prior	-0.006 [-0.014 to 0.001]	85% (-)	0.226 [0.030 to 0.426]	90% (+)
Neutral prior	-0.008 [-0.016 to 0.000]	90% (-)	0.237 [0.055 to 0.435]	90% (+)
Enthusiastic prior	-0.011 [-0.018 to -0.005]	97.5% (-)	0.305 [0.172 to 0.442]	97.5% (+)

3.4 DISCUSSION

While the survival of pediatric patients after cardiac surgery has improved over the last decades [28], a significant proportion of them still suffers from neurocognitive impairments [6]. As a result, vast attention has been given to identifying those factors, related to or independent from the clinical practice, which could affect the neurocognitive development of these children. We used Bayesian statistics to analyze the association between $SctO_2$ desaturation and mean $SctO_2$ and total IQ, 2-years after PICU admission.

In this study, increased dose of $SctO_2$ desaturation and lower mean $SctO_2$ in the first 12 and 24 hours of monitoring, independently increased the probability of having a lower total IQ 2-years after PICU admission. These findings are in line with previous studies that showed high vulnerability of the immature brain to prolonged periods of cerebral desaturation [2, 3, 7, 13, 29–31].

Validated critical thresholds for harmful cerebral $SctO_2$ are lacking. In literature desaturation is heterogeneously defined, with thresholds that range from 45% to 65% [29, 30, 32–34]. Although relatively high, the 65% threshold used in this study permitted to identify periods of desaturation of which dose independently increased the chances of lower total IQ at 2-years follow-up. Sensitivity analyses and interaction effect analysis confirmed the robustness of the results. The results may indicate that impaired brain oxygenation already occurs at higher oxygen saturation thresholds.

Adoption of skeptical, neutral and enthusiastic priors had almost no effect on the posterior β estimates, but it rather affected the level of certainty of the estimates, as shown by the standard deviations of the probability distributions in Figure 3.3. Hence, the association between $SctO_2$ predictors and total IQ not only results when an a priori association is hypothesized (enthusiastic priors) but also when an a priori lack of association (neutral priors) or a negative association (skeptical priors) is hypothesized.

The corresponding frequentist approach did not provide statistically significant results, this is likely due to a lack of statistical power. However, it is worth to observe that: first, Bayesian and frequentist analyses provide comparable β estimates values. Second, the credible interval resulting from the Bayesian model is always included in the 95% confidence interval resulting from the frequentist model.

Due to the limited size of the dataset and the data distribution, this study uniquely focused on the deleterious effects of desaturation on the developing brain. However, no observations can be made on the harmfulness of hyperoxia, of which long-term consequences should be further investigated in future studies.

It remains unclear whether $SctO_2$ desaturation and lower mean $SctO_2$ are similarly associated with neurocognitive outcomes when evaluated at later times points, for instance at 4-years or 10-years follow-up. In addition, although these results have no therapeutic implications, they may represent a foundation for future randomized clinical trials (RCT). The target of a possible RCT may be the maintenance of $SctO_2$ between 65% and the current $SctO_2$ upper range supported by literature (around 85%).

This study has some limitations. First, due to the single-center design of the data set, the results may not be generalizable and require external validation in a larger prospective cohort. Second, this is an observational study of prospectively collected data. Hence, only associations, and not causality, are shown by the results. Third, the $SctO_2$ was quantified with NIRS, which monitors the oxygen saturation of a regional area of the frontal lobe. In addition, the trace was obtained by averaging recordings from the right and left electrode. Therefore, considerations on the perfusion status of other areas of the brain outside the NIRS optical field or of one specific hemisphere cannot be made. Fourth, in this study we only focus on the cerebral oximetry during the first 12 and 24 hours after surgery, but brain damage could have been occurred at other moments in time (preoperatively, perioperatively or late postoperatively) and in other areas of the brain. Fifth, readings of the NIRS oximeter may vary among brands. Sixth, clinical or pathological causes of potentially harmful doses of $SctO_2$ desaturation were not investigated, but they should be explored in a future randomized prospective study. Seventh, given the reduced size of the dataset, desaturation was defined by using a cut off of 65% also for children with cyanosis, for whom a lower threshold might be more appropriate. Eight, the multivariable linear regression model was not adjusted for the duration of PICU stay. We acknowledge that the duration of PICU stay has been previously associated with poor neurodevelopment and hence it might play a role in explaining the long-term outcomes of children with CHD. However, it was excluded from the list of adjusting factors given the limited size of the dataset and given that it resulted highly correlated with the PIM3 score. Last, although multiple confounders were taken into account when

performing the analyses, and although sensitivity analyses on additional adjusting factors provided consistent results, we cannot exclude the eventuality of other factors that affect outcomes but were not included in the analysis.

Despite these limitations, this represents the first study that analyzes the association between $SctO_2$, measured by NIRS, and total IQ at 2-years follow-up in a large cohort of pediatric patients after cardiac surgery. A strength of this study is the exclusion of treatment bias, as the NIRS monitor was blinded to clinicians. Moreover, given the prospective design, the study could count on detailed data collection.

3.5 CONCLUSION

In critically ill children admitted to the PICU after congenital heart surgery, increased dose of desaturation, defined as the combination of time and intensity of $SctO_2<65\%$, and lower mean $SctO_2$ in the first 12 hours and 24 hours of monitoring time independently increased the probability of a lower total IQ at 2-years follow-up. Whether interventions to prevent desaturation and reduced mean $SctO_2$ would result in improved outcomes remains to be investigated.

ACKNOWLEDGMENT AND PERSONAL CONTRIBUTION

Study concept and design	Meyfroidt, Vlasselaers, Desmet, Güiza, Verstraete, Flechet, Van den Berghe, Vanhorebeek
Data acquisition	Van Cleemput, Jacobs, Wouters, Flechet
Statistical analysis	**Carra**
Interpretation of results	**Carra,** Meyfroidt, Güiza
Drafting of the manuscript	**Carra**
Manuscript revision	**All authors**
Principal investigator	Meyfroidt

3.A APPENDIX

3.A.1 *Cardiac surgery procedures amenable for inclusion*

Cardiac surgery procedures which were considered amenable for inclusion are reported below:

- Norwood procedure
- Glenn procedure
- Damus–Kaye–Stansel procedure
- Systemic-to-pulmonary artery shunt
- Pulmonary artery banding
- Atrial Septal Defect closure
- Patent Foramen Ovale repair or closure
- Widening of the supravalvular aorta
- Arterial switch operation
- Tricuspid valve repair
- Mitral Valve Repair
- Patch closure of sinus venosus
- Coarctatio of the aorta
- Surgical correction for congenital heart defects such as: vascular ring, ventricular septal defect, pulmonary venous connection, double-chambered right ventricle, troncus arteriosus type 1, total anomalous pulmonary venous return, tetralogy of Fallot, interrupted aortic arch, univentricular heart, Fontan procedure, Subaortic stenosis

3.A.2 *Definition of educational level of parents*

The education level is the average of the paternal and maternal educational level, and calculated based upon the 3-point scale subdivisions as made by the Algemene Directie Statistiek (Belgium; statbel.fgov.be/nl/) and the Centraal Bureau voor de Statistiek (The Netherlands; statline.cbs.nl): Low (=1), middle (=2) and high (=3) educational level.

3.A.3 *Definition of syndrome*

The presence of a syndrome was defined as an a priori clinical condition that could affect neurocognitive development [18]. It included all adverse clinical conditions that fall under the following categories:

- Genetically confirmed syndrome or pathogenic chromosomal abnormality
- Clearly defined syndrome, association or malformation without (identified) genetic aberration
- Polymalformative syndrome of unknown etiology
- Clear auditory or visual impairment without specified syndrome
- Congenital hypothyroidism due to thyroid agenesis
- Brain tumour or tumour with intracranial metastatic disease
- Pedopsychiatric disorder (e.g. autism spectrum disorder, (treatment for) attention deficit hyperactivity disorder)
- Severe medical disorder, not primarily neurologic, but suspected to alter psychomotor and/or mental performance
- Severe neonatal problem (e.g. severe asphyxia)
- Severe craniocerebral trauma or near-drowning
- Severe infectious encephalitis or drug-induced encephalopathy
- Infectious meningitis, encephalitis or Guillain-Barré
- Resuscitation and/or need for extracorporeal membrane oxygenation prior to inclusion
- Severe convulsions or stroke prior to inclusion

3.A.4 *Definition of early and late parenteral nutrition*

Early parenteral nutrition (PN) implied initiation of supplemental parenteral nutrition within the first 24 hours after admission to the PICU when enteral nutrition alone was insufficient to reach the caloric target. Late parenteral nutrition is defined as the withholding of parenteral nutrition in the first week of intensive care when enteral nutrition was insufficient to reach the caloric target. Although late parenteral nutrition might imply lower caloric intake, it has been shown to be clinically superior as compared with early parenteral nutrition in terms of incidence of new infections, time to recovery and long-term neurocognitive outcomes [18, 35].

3.A.5 *Description of the medical assessment*

Neurocognitive outcomes of part of the included children were assessed in the context of the PEPaNIC study. In the PEPaNIC study, a follow-up visit was performed by trained clinicians and psychologists and scheduled 2 years after the admission of the child at the PICU.

For this study, only the total IQ was analysed as indicator of long-term neurocognitive outcomes. The remaining children (children that were not originally part of the PEPaNIC study) were contacted by phone to schedule the follow-up specifically for this study. Follow-up protocol strictly followed the neurocognitive assessment protocol of the PEPaNIC study.

3.A.6 *Description of the total IQ tests*

General intellectual abilities were assessed with use of age-appropriate versions of the Wechsler Intelligence Quotient (IQ) tests. In detail, the Wechsler Preschool and Primary Scale of Intelligence (WPPSI-III-NL)[19] was used for children aged 2·5 years – 5 years 11 months (one version for age range 2 years 6 months – 3 years 11 months, and another version for age range 4 years – 5 years 11 months), the Wechsler Intelligence Scale for Children (WISC-III-NL) [20] was used for older children.

3.A.7 *Description of the method of calculation of dose of desaturation*

The total dose of desaturation experienced by each child was computed as follows:

$$total\ dose\ of\ SctO_2\ desaturation = \int_0^{tmax} f(s(t))$$

$$f(x) = \begin{cases} threshold - s(t) & s(t) < threshold \\ 0 & s(t) \geq threshold \end{cases}$$

Where:

- threshold: 65% (desaturation threshold)
- s: $SctO_2$ signal, s(t) refers to the value of the signal at time t

- tmax: 720 minutes (12 hours) or 1440 minutes (24 hours) , it refers to the maximum duration of the window from which the total dose of $SctO_2$ desaturation is extracted

3.A.8 *Construction of the Bayesian model*

The Bayesian model was built by using the No-U-Turn Sampler (NUTS). All non-informative priors were given a normal distribution, with a mean equal to 0 and a SD equal to 2 times the SD of the variable.

Informative priors were set for the variables syndrome and cyanotic cardiopathy. In detail, we assumed that the presence of syndrome would worsen the neurocognitive outcomes of the child by 5% [21–25, 36, 37]. Similarly, we assumed that the presence of cyanotic cardiopathy would worsen the neurocognitive outcomes of the child by 1% [26, 27]. Given that the mean total IQ at 2-years follow-up of children admitted in the PICU is 90.6 [18], the mean value and standard deviation of the priors set for of the syndrome and cyanotic cardiopathy are reported in Table 3.A.2.

Table 3.A.2 Parameters of the prior distributions for syndrome and cyanotic cardiopathy

COVARIATE	MEAN VALUE	SD
Syndrome	-4.53	10
Cyanotic cardiopathy	-0.91	10

3.A.9 *Examples of SctO2 sequences in the first 12 hours of PICU stay*

Examples of $SctO_2$ sequences in the first 12 hours of PICU stay, see Figure 3.A.1. For each signal, it is indicated the average loss of IQ points that is associated with the signal desaturation.

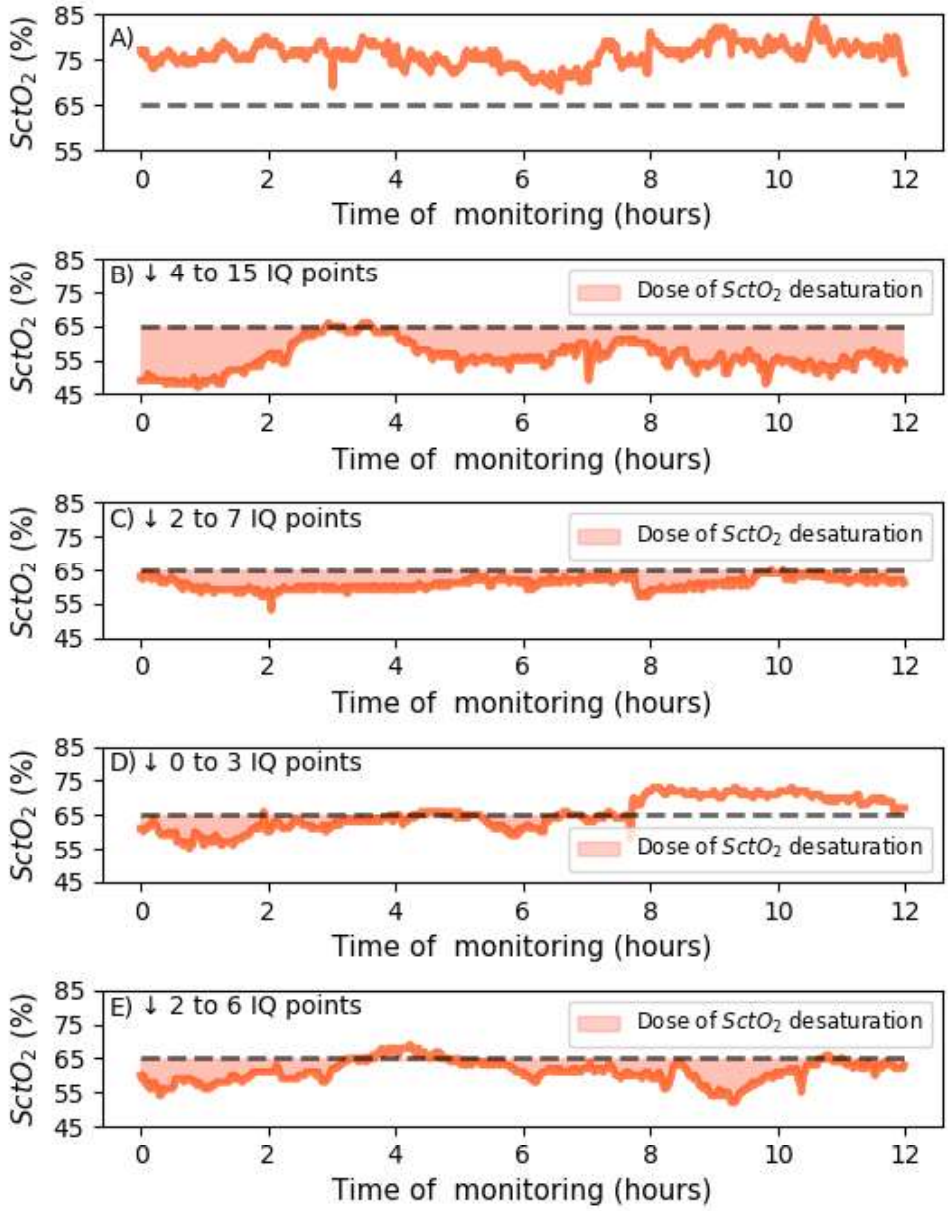


Figure 3.A.1 Sequences of $SctO_2$ recordings with different doses of desaturation, first 12 hours of monitoring time. The desaturation dose below 65% (unit of measurement: %·minutes) is represented by the red area between the signal and the grey dashed line which indicates the 65% desaturation threshold. Signal A) desaturation dose: 0%·minutes. Signal B) desaturation dose: 917 %·minutes. Signal

C) desaturation dose: 413 %·minutes. Signal D) desaturation dose: 189 %·minutes. Signal E) desaturation dose: 409 %·minutes. For each signal, the average drop of IQ points associated with its $SctO_2$ desaturation is indicated in respect with signal A), which shows no desaturation. For example, the desaturation dose of signal B) is associated with a loss of 4 to 15 IQ points than signal A). Similarly, the desaturation dose of signal C) is associated with a loss of 2 to 7 IQ points than signal A).

3.A.10 *Posterior β estimates of all co-variates and representation of the prior and posterior distributions*

Results of the Bayesian linear regression model are reported for all co-variates in terms of β estimates [80% credible interval] and probability of strictly negative/positive prior. For simplicity, and given that Bayesian analysis using neutral, skeptical and enthusiastic priors provided consistent results, only the results of the neutral prior, for desaturation and mean $SctO_2$ of the first 12 hours of monitoring time are reported, see Table 3.A.2 and Table 3.A.3 respectively. Visual representation of the prior and posterior distributions of all covariates is provided uniquely for the Bayesian model on mean $SctO_2$, Figure 3.A.2 to Figure 3.A.8.

Table 3.A.2 Posterior β estimates for all co-variates. Bayesian linear regression model for $SctO_2$ desaturation

Covariates	Posterior β estimates [80% credible interval]	Probability that the posterior β estimate is strictly negative(-)/positive(+)
Desaturation	-0.009 [-0.02 to -0.001]	90% (-)
Age	-0.04 [-0.11 to 0.02]	75% (-)
Syndrome	-8.73 [-15.24 to -3.15]	95% (-)
Nutrition strategy§	-5.56 [-9.07 to -2.06]	95% (-)
Cyanotic cardiopathy	-3.26 [-7.52 to 1.43]	80% (-)
PIM3	-1.42 [-3.61 to 0.73]	70% (-)
Intercept	93.00 [1.00 to 85.00]	97.5% (+)

§ Late PN vs early PN

Table 3.A.3 Posterior β estimates for all co-variates. Bayesian linear regression model for mean $SctO_2$

Covariates	Posterior β estimates [80% credible interval]	Probability that the posterior β estimate is strictly negative(-)/positive(+)
Mean $SctO_2$	0.23 [0.047 to 0.41]	90% (+)
Age	-0.04 [-0.10 to 0.02]	75% (-)
Syndrome	-9.08 [-15.32 to -3.23]	95% (-)
Nutrition strategy§	-5.32 [-8.98 to -1.72]	90% (+)
Cyanotic cardiopathy	-3.56 [-7.93 to 0.58]	85% (-)
PIM3	-1.39 [-3.58 to 0.86]	80% (-)
Intercept	77.35 [62.37 to 91.92]	97.5% (+)

§ Late PN vs early PN

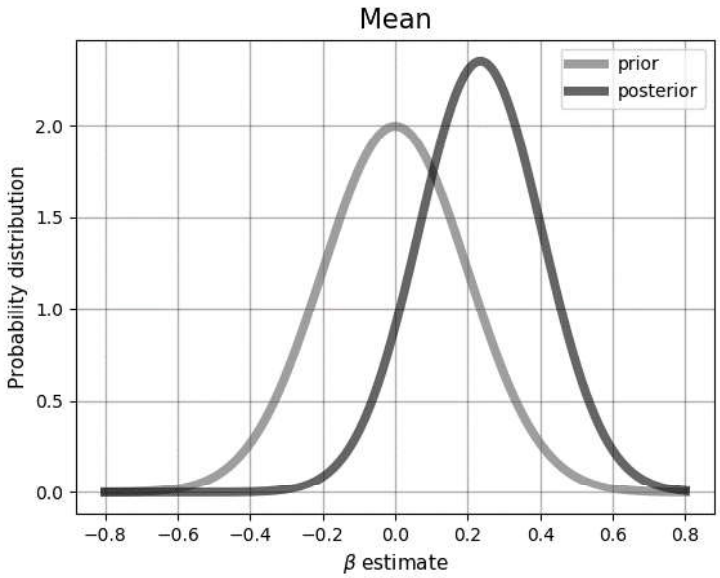


Figure 3.A.2 Comparison between the β priors and β posteriors probability distribution for the covariate mean $SctO_2$ of the multivariable Bayesian model on the relation between mean $SctO_2$ and total IQ (neutral prior, first 12 hours of monitoring time).

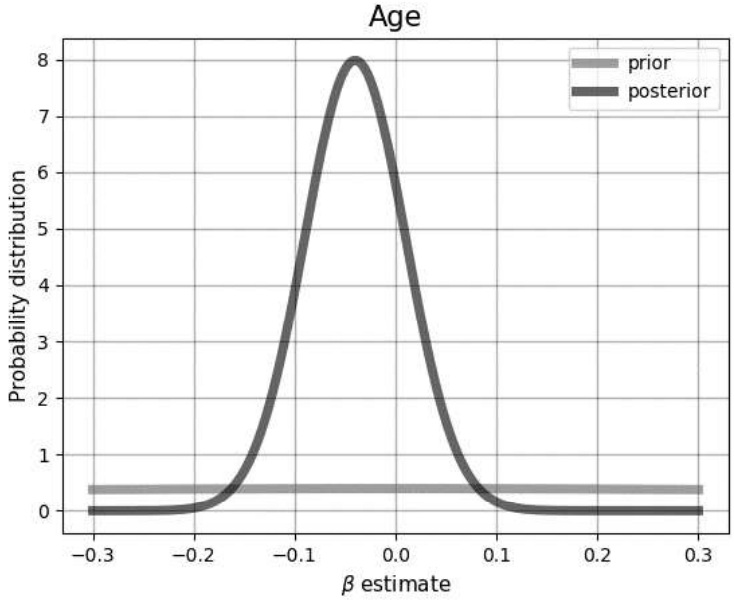


Figure 3.A.3 Comparison between the β priors and β posteriors probability distribution for the covariate age of the multivariable Bayesian model on the relation between mean $SctO_2$ and total IQ (neutral prior, first 12 hours of monitoring time). The β prior is non informative, as it is shown by the flat probability distribution.

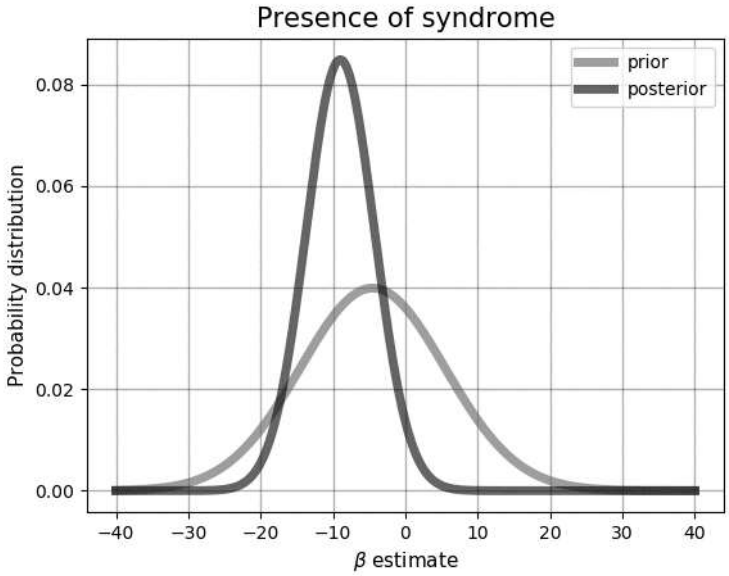


Figure 3.A.4 Comparison between the β priors and β posteriors probability distribution for the covariate presence of syndrome of the multivariable Bayesian model on the relation between mean $SctO_2$ and total IQ (neutral prior, first 12 hours of monitoring time).

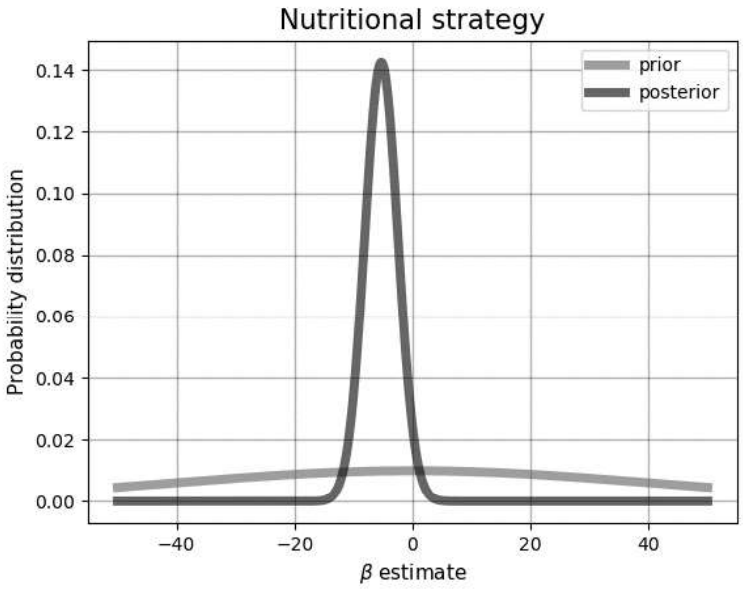


Figure 3.A.5 Comparison between the β priors and β posteriors probability distribution for the covariate nutritional strategy of the multivariable Bayesian model on the relation between mean $SctO_2$ and total IQ (neutral prior, first 12 hours of monitoring time).

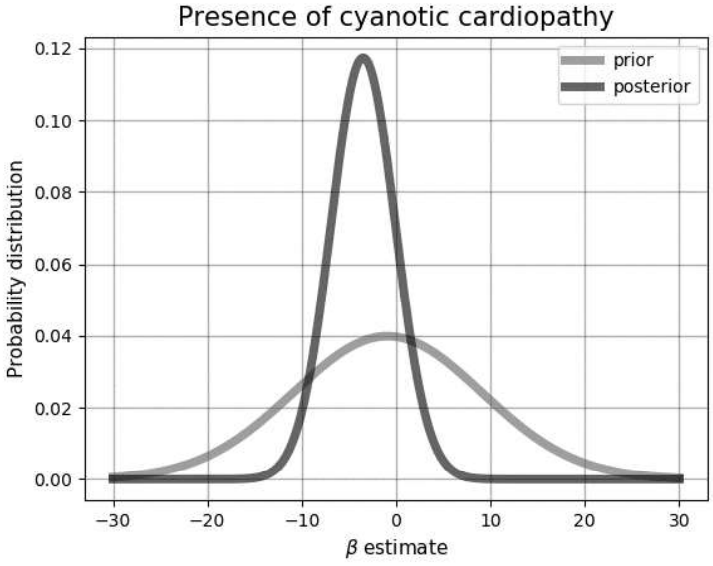


Figure 3.A.6 Comparison between the β priors and β posteriors probability distribution for the covariate presence of cyanotic cardiopathy of the multivariable Bayesian model on the relation between mean $SctO_2$ and total IQ (neutral prior, first 12 hours of monitoring time).

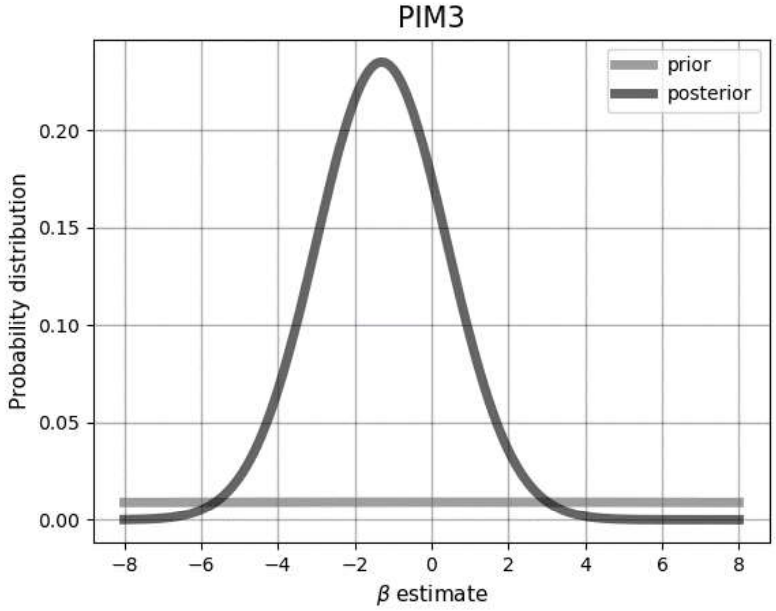


Figure 3.A.7 Comparison between the β priors and β posteriors probability distribution for the covariate nutritional strategy of the multivariable Bayesian model on the relation between mean SctO2 and total IQ (neutral prior, first 12 hours of monitoring time).

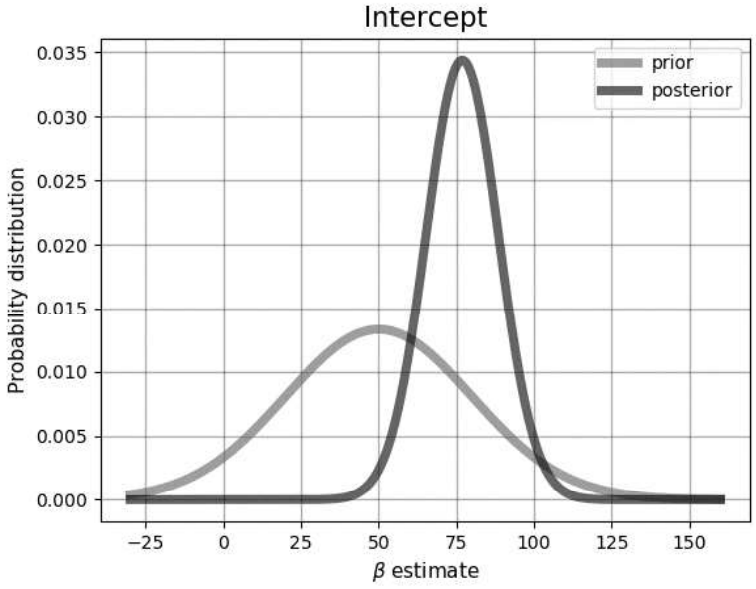


Figure 3.A.8 Comparison between the β priors and β posteriors probability distribution for the intercept of the multivariable Bayesian model on the relation between mean SctO2 and total IQ (neutral prior, first 12 hours of monitoring time).

3.A.11 *Sensitivity analysis on the definition of desaturation*

Given that Bayesian analyses using neutral, skeptical and enthusiastic priors provided consistent results, sensitivity analyses for different definitions of desaturation were performed only with the neutral prior. Desaturation was defined as $SctO_2$ below the thresholds of 60% and 55%. The results of the sensitivity analyses confirmed the robustness of the results of the main analysis, see Table 3.A.4 and Table 3.A.5 for desaturation defined as $SctO_2$ below 60% and 55% respectively.

Table 3.A.4 Results of a sensitivity analysis with desaturation defined as $SctO_2$ below 60%

	DESATURATION DOSE ($SctO_2$ < 60%)	
	Posterior β estimates [80% credible interval]	Probability that the posterior β estimate is strictly negative(-)/positive(+)
First 12-hours of $SctO_2$ monitoring	-0.09 [-0.16 to -0.03]	95% (-)
First 24-hours of $SctO_2$ monitoring	-0.02 [-0.067 to 0.025]	85% (-)

Table 3.A.5 Results of a sensitivity analysis with desaturation defined as $SctO_2$ below 55%

	DESATURATION DOSE ($SctO_2$ < 55%)	
	Posterior β estimates [80% credible interval]	Probability that the posterior β estimate is strictly negative(-)/positive(+)
First 12-hours of $SctO_2$ monitoring	-0.052 [-0.09 to -0.01]	95% (-)
First 24-hours of $SctO_2$ monitoring	-0.033 [-0.068 to 0.003]	85% (-)

3.A.12 *Sensitivity analyses on the effect of additional adjusting factors*

Sensitivity analyses were performed on the effect of additional adjusting factors on the results of the multivariable linear regression model. The multivariable linear regression model of the main analysis was additionally adjusted for the educational status of the parents and, in a separate analysis, for the duration of the cardiopulmonary bypass. The analyses were performed in two subsets of patients that had the required information, namely, in a subset of 76 patients for the additional adjusting factor "educational status of the parents" and a subset of 71 patients for the adjusting factor "duration of cardiopulmonary bypass". The results remained consistent with the main analysis. In particular, Table 3.A.6 reports the results of the linear regression model further adjusted for the educational status of the parents, while Table 3.A.7 reports the results of the linear regression model further adjusted for the duration of cardiopulmonary bypass.

Table 3.A.6 Multivariable linear regression analysis for the association between SctO2 features and total IQ. Additional adjusting factor: educational status parents. The analysis was performed on 76 children

	DESATURATION DOSE		MEAN $SctO_2$	
	Posterior β estimates [80% credible interval]	Probability that the posterior β estimate is strictly negative(-)/positive(+)	Posterior β estimates [80% credible interval]	Probability that the posterior β estimate is strictly negative(-)/positive(+)
First 12-hours of $SctO_2$ monitoring				
Skeptical prior	-0.009 [-0.013 to -0.004]	90% (-)	0.224 [0.101 to 0.358]	90% (+)
Neutral prior	-0.011 [-0.016 to -0.006]	95% (-)	0.248 [0.125 to 0.372]	95% (+)
Enthusiastic prior	-0.013 [-0.017 to -0.009]	97.5% (+)	0.308 [0.220 to 0.396]	97.5% (+)
First 24-hours of $SctO_2$ monitoring				
Skeptical prior	-0.009 [-0.014 to -0.004]	90% (-)	0.254 [0.133 to 0.393]	90% (+)
Neutral prior	-0.011 [-0.017 to -0.007]	95% (-)	0.268 [0.147 to 0.389]	95% (+)
Enthusiastic prior	-0.013 [-0.17 to -0.009]	97.5% (-)	0.317 [0.233 to 0.402]	97.5% (+)

Table 3.A.7 Multivariable linear regression analysis for the association between SctO2 features and total IQ. Additional adjusting factor: duration of CPB. The analysis was performed on 71 children

	DESATURATION DOSE		MEAN $SctO_2$	
	Posterior β estimates [80% credible interval]	Probability that the posterior β estimate is strictly negative(-)/positive(+)	Posterior β estimates [80% credible interval]	Probability that the posterior β estimate is strictly negative(-)/positive(+)
First 12-hours of $SctO_2$ monitoring				
Skeptical prior	-0.005 [-0.010 to 0.000]	80% (-)	0.169 [0.029 to 0.311]	85% (+)
Neutral prior	-0.007 [-0.013 to -0.002]	85% (-)	0.199 [0.074 to 0.332]	90% (+)
Enthusiastic prior	-0.011 [-0.016 to -0.007]	97.5% (-)	0.287 [0.200 to 0.406]	97.5% (+)
First 24-hours of $SctO_2$ monitoring				
Skeptical prior	-0.003[-0.009 to 0.003]	70% (-)	0.146 [0.006 to 0.291]	80% (+)
Neutral prior	-0.005 [-0.012 to 0.000]	80% (-)	0.179 [0.035 to 0.291]	85% (+)
Enthusiastic prior	-0.010[-0.015 to -0.050]	95% (-)	0.276 [0.188 to 0.373]	97.5% (+)

3.A.13 *Interaction analysis*

No association was found between presence of cyanosis after surgery and dose of desaturation (p=0.47 and p=0.22 respectively for the first 12 hours and 24 hours of monitoring). Presence of cyanosis after surgery was associated with lower mean $SctO_2$ during the first 12 hours and 24 hours of monitoring time, *p-values* of respectively $p < 0.001$ and $p < 0.001$. However, no interaction effect was found between the presence of cyanosis after surgery and mean $SctO_2$ in relation with long-term total IQ. Results of the β estimates for the neutral prior are reported in Table 3.A.8 (first 12 hours).

Table 3.A.8 Results of the interaction term analysis for mean $SctO_2$ and total IQ

Covariates	Posterior β estimates [80% credible interval]	Probability that the posterior β estimate is strictly negative(-)/positive(+)
Mean $SctO_2$	0.23 [0.13 to 0.36]	90% (+)
Age	-0.04 [-0.08 to 0.01]	75% (-)
Syndrome	-8.92 [-12.86 to -4.75]	95% (-)
Nutrition strategy	-5.33 [-7.68 to -2.95]	95% (-)
Cyanosis after surgery	-3.31 [-11.6 to 4.15]	65% (-)
PIM3	-1.36 [-2.76 to 0.09]	75% (-)
Mean $SctO_2$ * cyanosis after surgery	-0.003 [-0.121 to 0.122]	50% (-)

BIBLIOGRAPHY

1. Hoffman JI., Kaplan S (2002) The incidence of congenital heart disease. J Am Coll Cardiol 39:1890–1900. https://doi.org/10.1016/S0735-1097(02)01886-7

2. Glass TJA, Seed M, Chau V (2019) Chapter 15 - Congenital Heart Disease: An Important Cause of Brain Injury and Dysmaturation, Third Edit. Elsevier Inc.

3. Galli KK, Zimmerman RA, Jarvik GP, et al (2004) Periventricular leukomalacia is common after neonatal cardiac surgery. J Thorac Cardiovasc Surg 127:692–704. https://doi.org/10.1016/j.jtcvs.2003.09.053

4. Bellinger DC, Wypij D, Kuban KCK, et al (1999) Developmental and Neurological Status of Children at 4 Years of Age After Heart Surgery With Hypothermic Circulatory Arrest or Low-Flow Cardiopulmonary Bypass. Circulation 100:526–532. https://doi.org/10.1161/01.CIR.100.5.526

5. Majnemer A, Limperopoulos C (1999) Developmental progress of children with congenital heart defects requiring open heart surgery. Semin Pediatr Neurol 6:12–9

6. Gaynor JW, Stopp C, Wypij D, et al (2015) Neurodevelopmental Outcomes After Cardiac Surgery in Infancy. Pediatrics 135:816–825. https://doi.org/10.1542/peds.2014-3825

7. Kinney HC, Panigrahy A, Newburger JW, et al (2005) Hypoxic-ischemic brain injury in infants with congenital heart disease dying after cardiac surgery. Acta Neuropathol 110:563–578. https://doi.org/10.1007/s00401-005-1077-6

8. Wernovsky G, Licht DJ (2016) Neurodevelopmental Outcomes in Children With Congenital Heart Disease—What Can We Impact? Pediatr Crit Care Med 17:S232–S242. https://doi.org/10.1097/PCC.0000000000000800

9. Ghanayem NS, Wernovsky G, Hoffman GM (2011) Near-infrared spectroscopy as a hemodynamic monitor in critical illness. Pediatr Crit Care Med 12:S27–S32. https://doi.org/10.1097/PCC.0b013e318221173a

10. Zulueta JL, Vida VL, Perisinotto E, et al (2013) The Role of Intraoperative Regional Oxygen Saturation Using Near Infrared Spectroscopy in the Prediction of Low Output Syndrome After Pediatric Heart Surgery. J Card Surg 28:446–452.

https://doi.org/10.1111/jocs.12122

11. Fenton KN, Freeman K, Glogowski K, et al (2005) The significance of baseline cerebral oxygen saturation in children undergoing congenital heart surgery. Am J Surg 190:260–263. https://doi.org/10.1016/j.amjsurg.2005.05.023

12. Dent CL, Spaeth JP, Jones B V., et al (2005) Brain magnetic resonance imaging abnormalities after the Norwood procedure using regional cerebral perfusion. J Thorac Cardiovasc Surg 130:1523–1530. https://doi.org/10.1016/j.jtcvs.2005.07.051

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

14. Zaleski KL, Kussman BD (2020) Near-Infrared Spectroscopy in Pediatric Congenital Heart Disease. J Cardiothorac Vasc Anesth 34:489–500. https://doi.org/10.1053/j.jvca.2019.08.048

15. Vida VL, Tessari C, Cristante A, et al (2016) The Role of Regional Oxygen Saturation Using Near-Infrared Spectroscopy and Blood Lactate Levels as Early Predictors of Outcome After Pediatric Cardiac Surgery. Can J Cardiol 32:970–977. https://doi.org/10.1016/j.cjca.2015.09.024

16. Slater JP, Guarino T, Stack J, et al (2009) Cerebral Oxygen Desaturation Predicts Cognitive Decline and Longer Hospital Stay After Cardiac Surgery. Ann Thorac Surg 87:36–45. https://doi.org/10.1016/j.athoracsur.2008.08.070

17. Flechet M, Güiza F, Scharlaeken I, et al (2019) Near-Infrared–Based Cerebral Oximetry for Prediction of Severe Acute Kidney Injury in Critically Ill Children After Cardiac Surgery. Crit Care Explor 1:e0063. https://doi.org/10.1097/CCE.0000000000000063

18. Verstraete S, Verbruggen SC, Hordijk JA, et al (2018) Long-term developmental effects of withholding parenteral nutrition for 1 week in the paediatric intensive care unit: a 2-year follow-up of the PEPaNIC international, randomised, controlled trial. Lancet Respir Med 7:141–153. https://doi.org/10.1016/S2213-2600(18)30334-5

19. Hendriksen J, Hurks P (2010) WPPSI-III-NL | Wechsler Preschool and Primary Scale of Intelligence

20. D. Wechsler (2005) WISC-III-NL | Wechsler Intelligence Scale for Children-III. Amsterdam

21. Miatton M, De Wolf D, François K, et al (2006) Neurocognitive consequences of surgically corrected congenital heart defects: A review. Neuropsychol Rev 16:65–85. https://doi.org/10.1007/s11065-006-9005-7

22. Palmer SL, Hassall T, Evankovich K, et al (2010) Neurocognitive outcome 12 months following cerebellar mutism syndrome in pediatric patients with medulloblastoma. Neuro Oncol. https://doi.org/10.1093/neuonc/noq094

23. Virtanen R, Korhonen T, Fagerholm J, Viljanto J (1999) Neurocognitive Sequelae of Scaphocephaly. Pediatrics 103:791–795. https://doi.org/10.1542/peds.103.4.791

24. van der Sluijs Veer L, Kempers MJE, Wiedijk BM, et al (2012) Evaluation of Cognitive and Motor Development in Toddlers With Congenital Hypothyroidism Diagnosed by Neonatal Screening. J Dev Behav Pediatr 33:633–640. https://doi.org/10.1097/DBP.0b013e3182690727

25. Armstrong-Wells J, Bernard TJ, Boada R, Manco-Johnson M (2010) Neurocognitive outcomes following neonatal encephalopathy. NeuroRehabilitation 26:27–33. https://doi.org/10.3233/NRE-2010-0533

26. Hövels-Gürich HH, Seghaye M-C, Däbritz S, et al (1997) Cognitive and motor development in preschool and school-aged children after neonatal arterial switch operation. J Thorac Cardiovasc Surg 114:578–585. https://doi.org/10.1016/S0022-5223(97)70047-3

27. Limperopoulos C, Majnemer A, Shevell MI, et al (2002) Predictors of developmental disabilities after open heart surgery in young children with congenital heart defects. J Pediatr 141:51–58. https://doi.org/10.1067/mpd.2002.125227

28. Oster ME, Lee KA, Honein MA, et al (2013) Temporal Trends in Survival Among Infants With Critical Congenital Heart Defects. Pediatrics 131:e1502–e1508. https://doi.org/10.1542/peds.2012-3435

29. Kussman BD, Wypij D, Laussen PC, et al (2010) Relationship of intraoperative cerebral oxygen saturation to neurodevelopmental outcome and brain magnetic resonance imaging at 1 year of age in infants undergoing biventricular repair. Circulation 122:245–254.

https://doi.org/10.1161/CIRCULATIONAHA.109.902338

30. Spaeder MC, Klugman D, Skurow-Todd K, et al (2017) Perioperative Near-Infrared Spectroscopy Monitoring in Neonates with Congenital Heart Disease: Relationship of Cerebral Tissue Oxygenation Index Variability with Neurodevelopmental Outcome. Pediatr Crit Care Med 18:213–218. https://doi.org/10.1097/PCC.0000000000001056

31. Hoffman GM, Ghanayem NS, Scott JP, et al (2017) Postoperative Cerebral and Somatic Near-Infrared Spectroscopy Saturations and Outcome in Hypoplastic Left Heart Syndrome. Ann Thorac Surg 103:1527–1535. https://doi.org/10.1016/j.athoracsur.2016.09.100

32. McNeill S, Gatenby JC, McElroy S, Engelhardt B (2011) Normal cerebral, renal and abdominal regional oxygen saturations using near-infrared spectroscopy in preterm infants. J Perinatol 31:51–57. https://doi.org/10.1038/jp.2010.71

33. Dix LML, van Bel F, Lemmers PMA (2018) Monitoring Cerebral Oxygenation in Neonates: An Update. Front Pediatr 5:1–9. https://doi.org/10.3389/fped.2017.00160

34. Neshat Vahid S, Panisello JM (2014) The state of affairs of neurologic monitoring by near-infrared spectroscopy in pediatric cardiac critical care. Curr Opin Pediatr 26:299–303. https://doi.org/10.1097/MOP.0000000000000098

35. Fivez T, Kerklaan D, Mesotten D, et al (2016) Early versus Late Parenteral Nutrition in Critically Ill Children. N Engl J Med 374:1111–1122. https://doi.org/10.1056/NEJMoa1514762

36. Mesotten D, Gielen M, Sterken C, et al (2012) Neurocognitive development of children 4 years after critical illness and treatment with tight glucose control: A randomized controlled trial. JAMA - J Am Med Assoc 308:1641–1650. https://doi.org/10.1001/jama.2012.12424

37. Nandi-Munshi D, Taplin CE (2015) Thyroid-Related Neurological Disorders and Complications in Children. Pediatr Neurol 52:373–382. https://doi.org/10.1016/j.pediatrneurol.2014.12.005

4

ASSOCIATION OF DOSE OF INTRACRANIAL HYPERTENSION WITH OUTCOME IN SUBARACHNOID HEMORRHAGE

ADAPTED FROM: **Carra G.** et al. Association of dose of intracranial hypertension with outcome in subarachnoid hemorrhage. Neurocritical Care. 2021 Jun. 34(3):722-730.

Presented as:

- Oral presentation at the 32nd European Society of Intensive Care Medicine (ESICM) Annual Congress, Berlin, Germany, 2019.
- Poster presentation at 10th EuroNeuro conference, Brussels, Belgium, 2018.
- Poster presentation at the 39th International Symposium on Intensive Care and Emergency Medicine (ISICEM), Brussels, Belgium, 2019.

ABSTRACT

PURPOSE: In patients with aneurysmal subarachnoid hemorrhage (aSAH) the burden of intracranial pressure (ICP) and its contribution to outcomes remains unclear. In this multicenter study, the independent association between intensity and duration, or "dose", of episodes of intracranial hypertension and 12-month neurological outcomes was investigated.

METHODS: Retrospective analysis of multicenter prospectively collected data of 98 adult patients with aSAH amendable to treatment. Patients were admitted to the Intensive Care Unit of two European centers (Medical University of Innsbruck (Austria) and San Gerardo University Hospital of Monza (Italy)), from 2009 to 2013. The dose of intracranial hypertension was visualized. The obtained visualizations allow to investigate the association between intensity and duration of episodes of intracranial hypertension and the 12-month neurological outcomes of the patients, assessed with the Glasgow Outcome Score. The independent association between the cumulative dose of intracranial hypertension and outcome for each patient, was investigated using multivariable logistic regression models, corrected for age, occurrence of delayed cerebral ischemia and Glasgow Coma Scale at admission.

RESULTS: The combination of duration and intensity defined the tolerance to intracranial hypertension for the 2 cohorts of patients. A semi-exponential transition divided ICP doses that were associated with better outcomes (in blue) with ICP doses associated with worse outcomes (in red). In addition, in both cohorts an independent association was found between the cumulative time that the patient experienced ICP doses in the red area and long-term neurological outcomes. The ICP pressure time burden was a stronger predictor of outcomes than the cumulative time spent by the patients with an ICP above 20 mmHg.

CONCLUSIONS: In two cohorts of patients with aSAH, an association between duration and intensity of episodes of elevated ICP and 12-month neurological outcomes could be demonstrated and was visualized in a color-coded plot.

4.1 INTRODUCTION

Aneurysmal subarachnoid haemorrhage (aSAH) is a form of stroke that is characterized by the extravasation of blood in the subarachnoid space due to the rupture of an aneurysm [1, 2].
In the first days after the initial aneurysm rupture, rebleeding is the most concerning risk. Hence, securing the ruptured aneurysm through coiling or clipping is done as early as possible. Once the aneurysm is secured, the management of the patient is aimed at the prevention or treatment of subsequent complications [2], of which delayed cerebral ischemia (DCI) is the most prevalent and feared [3]. Increased intracranial pressure (ICP) is also a main concern, since it may occur in up to 81% of patients [4, 5]. Management guidelines for the treatment of elevated ICP in aSAH are mainly based on the recommendations for traumatic brain injury (TBI). As in TBI, it is suggested to treat sustained ICP elevations above 20-22 mmHg [6, 7]. Guidelines refer to a fixed threshold, despite it was demonstrated that the concept of "dose", which combines intensity and duration of an ICP elevation, better quantifies ICP burden in patients with TBI [8, 9]. The association between dose of ICP and long-term neurological outcomes of patients with TBI was visualized with the method proposed by Güiza et al. [10]. For aSAH, such large epidemiological data on ICP thresholds are lacking [11, 12].

In the present study, we hypothesized that the concept of ICP dose applies to patients with aSAH. The association between ICP dose and outcomes was visualized by applying the method introduced by Güiza et al [10].

4.2 METHODS

4.2.1 *Dataset and study population*

This retrospective analysis was performed on two prospectively collected datasets. The first dataset, which was collected at the Department of Neurology, Neurocritical Care Unit of the Medical University of Innsbruck (Austria), included data of 46 consecutive patients admitted to the local ICU from 2010 to 2016. The second dataset included 52 consecutive patients admitted to the local ICU of the San Gerardo University Hospital of Monza (Italy) from 2009 to 2013 [5].

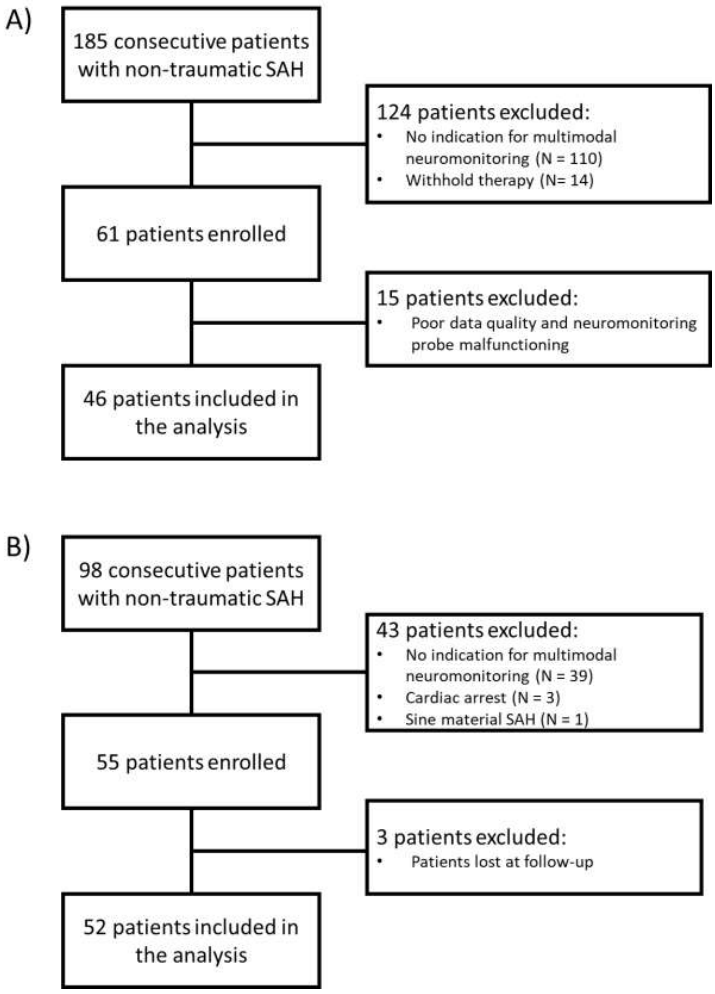


Figure 4.1 Study flow diagrams for the two cohorts: panel A) Innsbruck cohort panel B) Monza cohort.

Adult patients (age>18 years old) with aSAH amenable to treatment through surgical clipping or endovascular coiling and intraparenchymal ICP in place were included in the study. Study flow diagrams are reported in Figure 4.1. The policy of the Medical University of Innsbruck expects external ventricular drain (EVD) and ICP sensor placement in patients with poor clinical status (Hunt and Hess [H&H] score 4-5) at admission or neuroworsening in the first 24 hours, or clinical or radiological signs of potential increased intracranial pressure. In the San Gerardo University Hospital, EVD and ICP sensors are placed in patients deemed to require continuous cerebrospinal fluid (CSF) drainage [5]. The sensors are usually inserted when the patient has World Federation of Neurosurgical Societies (WFNS) ≥ 3 or poor Fisher score or presents early neuroworsening (within

the first 12 hours - 24 hours) after EVD placement. Patients with decompressive craniectomy or moribund at ICU admission were excluded from the study. The datasets include age, gender, modified Fisher (mFisher) score [13], H&H score [14], the WFNS score [15], the Glasgow Coma Scale (GCS) before sedation, continuous recordings of ICP and mean arterial blood pressure (MABP), indication and time of surgical therapy (coiling vs clipping), the amount of CSF drained during ICU stay and indications on the occurrence of DCI or re-bleeding. All illness severity scores were quantified and collected at admission in the ICU. DCI was defined as the development of new focal neurological signs or deterioration in level of consciousness, (decrease ≥ 2 points on the Glasgow Coma Scale [16] or neurological deterioration lasting more than 1 hour), or the appearance of new infarctions on computed tomography or magnetic resonance imaging not attributable to other causes [5]. All patients received sedatives and vasopressors. Information about their daily administration was included in the datasets as binary variables, detailed information about the dose or time of administration could not be retrieved. Clinical neurological outcomes were assessed by a trained neurologist and quantified with the 12-month Glasgow Outcome Score (GOS) [17].

Continuous ICP and MABP signals were measured and collected differently for the two centres. In the Innsbruck cohort, the samples were collected every second and the values stored in the server of the patient data management system (CentricityTM Critical Care, 8.1, SP7, GE, Boston, Massachusetts, USA) every 5 minutes. In the Monza cohort, the recordings were collected with the Innovian Solution Suite (Draeger Medical) that automatically samples continuous monitoring signals with a sampling frequency of 150Hz and averages the collected values across 1 minute.

To allow simultaneous ICP monitoring and CSF drainage, both centres used EVD as well as intraparenchymal ICP sensors (Innsbruck: Neurovent-P-temp, Raumedic®, Helmbrechts, Germany. Monza: Codman Microsensor® ICP Transducer, Johnson & Johnson Professional, Inc., USA). The clinical management of the patients followed the international American Heart Association (AHA) and Neurocritical Care Society (NCS) guidelines during the data collection [18, 19]. In both centres, EVD was kept open during the monitoring and ICP treated when its value rose above 20 mmHg for more than 10-20 minutes. Given that ICP measurements are unreliable when acquired from an open EVD, only data from the intraparenchymal probe were used for the analysis.

The study was approved by the ethics committee of both hospitals. For the Innsbruck cohort, written informed consent was obtained from the participants or their legal representatives. Need of informed consent was waived for the Monza cohort due to the retrospective nature of the analysis.

4.2.2 Statistical analysis of the datasets and visualization

Continuous and ordinal variables with non-normal distribution, such as age, mFisher score, GOS, H&H, WFNS and the GCS were analysed by computing median, interquartile range [IQR], 25th and 75th percentile, Chi-Square test for proportion, Mann Whitney U-test and Kolmogorov-Smirnov (KS) test. Categorical variables such as the occurrence of DCI or re-bleeding were analysed with the Chi-Square test for proportion. Statistical significance was set at a p value below 0.05. Effect size were computed with the Vargha and Delaney's A (A) [20] for continuous and ordinal variables and with odds ratios (OR) for categorical variables.
The Vargha and Delaney's A ranges from 0 and 1, it reports the probability that a value of one group is greater than the value of another group. As such, an A equal to 0.5 indicates that the two groups are stochastically equal, i.e. there is no effect. An A equal to 1 indicates that every value of the first group is higher than the values of the second group, and vice versa for an A equal to 0. Therefore, an A of 0 or 1 equally indicate a strong effect size.
The average percentage of monitoring time spent by each patient with an ICP above 20 mmHg and with a MABP above 80 mmHg were computed for the entire monitoring time and per day of monitoring.

The independent association between dose of ICP and neurological outcomes of patients with aSAH was investigated with the method of Güiza et al. [10]. It results in a 2D color-coded plot representing the correlation between the frequency of occurrence of ICP episodes of a certain dose (duration per intensity) and the severity of long-term neurological outcome, quantified with the 12-month GOS. Doses of ICP that occur more frequently in patients that present poor outcomes are represented in red, while doses of ICP which occur more frequently in patients with good outcomes are represented in blue. Doses of ICP that similarly occur in patients with good and poor long-term outcomes are represented in green and yellow (transition colors between red and blue). In the text, we will call "transition curves" the points of transition between regions associated with good GOS (blue region) and poor GOS (red region).

The 2-D color-coded plots were extracted for each center, using the entire cohort. To account for known confounders, the plots were re-extracted after excluding patients that had re-bleeding or DCI.

For each patient, we computed the cumulative percentage of monitoring time in which the patient had ICP doses in the red area of the color-coded plot, i.e. ICP doses associated with poor outcomes. This measure, which will be called "ICP pressure-time burden", was computed according to the transition curves of each cohort.

To further validate the visualization plots, and account for known correcting factors, the independent association between the ICP pressure-time burden and outcomes was evaluated with multivariable logistic regression models corrected for age, DCI and GCS. Outcomes were dichotomized and defined as good (GOS 4-5) versus poor (GOS 1-3) outcomes. The same multivariate models were built substituting the ICP pressure-time burden with the percentage of time with ICP above 20 mmHg. To account for variations in the monitoring time, the same analyses were performed by limiting the ICP recordings of the Innsbruck cohort to the first week of monitoring time.

All the analyses were performed separately for the two cohorts given the differences in data collection (prospective for the Innsbruck cohort and retrospective for the Monza cohort) and in sampling frequencies used by the two cohorts.

All the analyses were computed in Python 3.5.4 (https://www.python.org/) and Scipy 0.19.1 (https://www.scipy.org/).

4.3 RESULTS

Baseline clinical characteristics and outcomes are shown in Table 4.1. The two cohorts do not differ significantly in age, gender, mFisher score, WFNS score, GCS or H&H score.

The percentage of patients that underwent coiling or clipping was comparable in the two centers, despite a small effect size (p= 0.13 and p = 0.17 respectively, OR = 1.69). Importantly, in the Monza cohort the incidence of re-bleeding was significantly higher than in the Innsbruck cohort (p = 0.01, OR = 3.02). On the contrary, the incidence of DCI was comparable between the datasets (p = 0.30, OR = 0.77).

The average duration of the monitoring time with intraparenchymal devices was 14 days ± 7 days and 5 days ± 2 days for the Innsbruck and Monza

cohorts respectively (p<0.001, A = 0.93). During the first monitoring week, the patients experienced comparable percentages of cumulative time above 20 mmHg, equal to 2.4% and 1.6% (p = 0.38, A = 0.52) in Innsbruck and Monza, respectively; see Appendix 4.A.1 (Figure 4.A.1). However, the daily percentage of cumulative time above 20 mmHg, was significantly higher in Monza in the first day of monitoring (p = 0.04, A = 0.40) and significantly higher in Innsbruck for days 6 (p = 0.03, A = 0.61) and 7 (p = 0.01, A = 0.68); see Appendix 4.A.1 (Figure 4.A.2). Similarly, patients from Innsbruck experienced higher daily MABP values for the first 5 days of monitoring (p<0.001 and A > 0.9 for days 1 to 4, p = 0.002, A = 0.8 for day 5, and p = 0.37, A = 0.76 and p = 0.44, A = 0.58 for days 6 and 7 respectively); see Appendix 4.A.1 (Figure 4.A.3).

During the first week of ICU stay, the number of days that patients received sedatives was significantly higher (p<0.001, A = 0.77) in Innsbruck than in Monza with a median [IQR] of 7 [7-7] vs 6 [4-7]. Likewise, the number of days that patients received vasopressors was significantly higher (p = 0.008, A = 0.69) in Innsbruck than in Monza with a median [IQR] of 7 [6-7] vs 6 [4-7], see Figure 4.2.

Lastly, the average amount of CSF that was drained daily over the first week of ICU stay was higher in Innsbruck (p<0.001, A = 0.77). Major information on the comparison of the post-surgery clinical management of patients with aSAH for the centers can be found in Appendix 4.A.2. The dichotomized 12-month neurological outcomes were comparable (Chi-square test of proportion p = 0.53 and p = 0.79 for GOS 1,2,3 and GOS 4,5 respectively, OR = 0.87).

Table 4.1 Baseline clinical characteristics and outcomes

VARIABLE	INNSBRUCK	MONZA	p	EFFECT SIZE
Number of patients	46	52		
Age, median [IQR]	57 [47 – 67]	57 [50 – 65]	0.30	0.47[a]
Gender, females, n (%)	28 (61)	34 (65)	0.30	1.21[b]
mFisher score, median [IQR]	4 [3 – 4]	4 [4 – 4]	0.18 0.99[c]	0.46[a]
WFNS, median [IQR]	5 [4 – 5]	5 [4 – 5]	0.17 0.90 [c]	0.55[a]

H&H, median [IQR]	4 [3 – 5]	4 [4 – 5]	0.49 0.72[c]	0.49[a]
GCS, median [IQR]	9 [4-13]	7 [5-13]	0.25 0.21[c]	0.53[a]
Surgical intervention:				1.69[b]
Coiling, n (%)	13 (28)	22 (42)	0.13	
Clipping, n (%)	31 (67)	28 (54)	0.17	
None, n (%)	2 (5)	2 (4)	0.90	
Time between ICU admission and surgical intervention (coil/clipping), days, median [IQR]	0 [0 - 0]	0 [0 - 0]	0.47	0.50[a]
DCI, n (%)	17 (37)	16 (31)	0.30	0.77[b]
Re-bleeding, n (%)	5 (11)	14 (27)	0.01	3.02[b]
Duration of monitoring, days, mean (SD)	14 (7)	5 (2)	<0.001	0.93[a]
Time ICP>20 mmHg , (% of monitoring time)	2.4	1.6	0.38	0.52[a]
Duration of treatment with sedatives (first week), days, median [IQR]	7 [7-7]	7 [6-7]	<0.001	0.77[a]
Duration of treatment with vasopressors (first week), days, median [IQR]	7 [6-7]	6 [4-7]	0.008	0.69[a]
Average daily drained CSF, mL, mean (SD)	195 (69)	127 (57)	<0.001	0.77[a]
12-month GOS:				0.87[b]
GOS 1-2-3, n (%)	18 (39)	22 (42)	0.53	
GOS 4-5, n (%)	28 (61)	30 (58)	0.79	

[a] Effect size were computed with the Vargha and Delaney's *A* for continuous and ordinal variables. Common effect size indexes for Vargha and Delaney's *A* are: small = $0.56 < A < 0.64$ and $0.34 < A < 0.44$; medium = $0.64 < A < 0.71$ and $0.29 < A < 0.34$; large $A \geq 0.71$ and $A \leq 0.29$.

[b] Effect size were computes with odds ratios (*OR*) for categorical variables. Common effect size indexes for *ORs* are as follows: small=1.5, medium=2 and large=3.

[c] KS *p* is computed only for the ordinal variables with non-normal distribution.

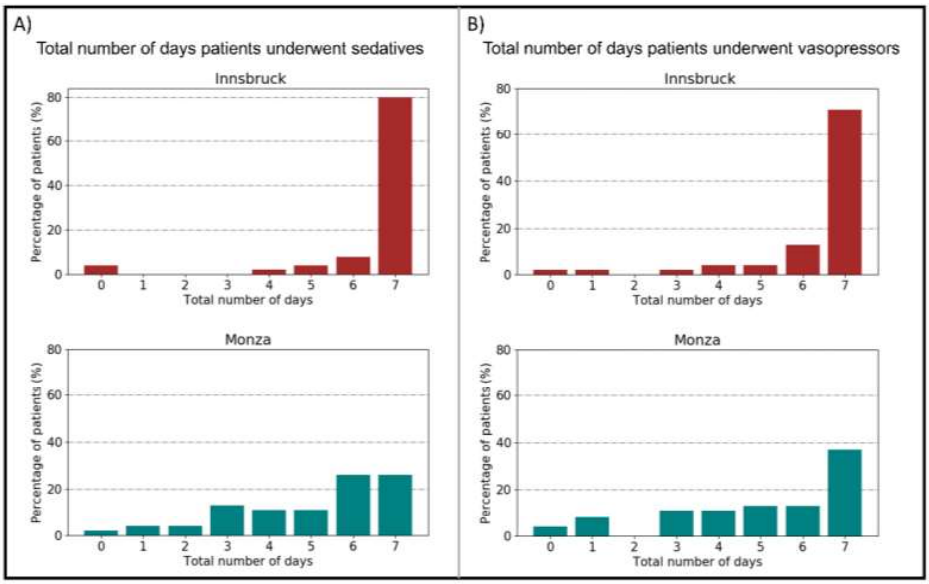


Figure 4.2 Bar plots of the total number of days patients from Innsbruck and Monza underwent a specific treatment. In detail: panel A) total number of days patients underwent sedatives, panel B) total number of days patients underwent vasopressors. Values are shown in terms of percentage.

4.3.1 *Visualization of ICP burden*

The color-coded graphs from the two datasets showed a time-intensity-dependent association between the dose of ICP and neurological outcomes at 12 months, as shown in Figure 4.3. The zones associated with good and poor outcomes were separated by a semi-exponential transition. The graph corresponding to the Innsbruck cohort showed a transition with a gradual exponential decrease from a dose of ICP of 10 mmHg for 350 minutes to 15 mmHg for 10 minutes. The transition for the Monza patients was shifted to the left, with a sharper decrease that started from about a dose of 5 mmHg for 160 minutes to 15 mmHg for 5 minutes. Patients from Monza did not experienced enough episodes of prolonged elevated ICP to compute part of the graph, which is indicated with a white region.

The time-intensity-dependent association between dose of ICP and 12-month GOS remained when excluding from the analysis patients with re-bleeding or DCI. Figure 4.4 shows the transition curves of the two centers when excluding, or not, patients with rebleeding or DCI from the analysis. The transition of the Monza cohort resulted shifted towards higher values than the one in Figure 4.3, extending from 5 mmHg for 340 minutes to 15 mmHg for 15 minutes.

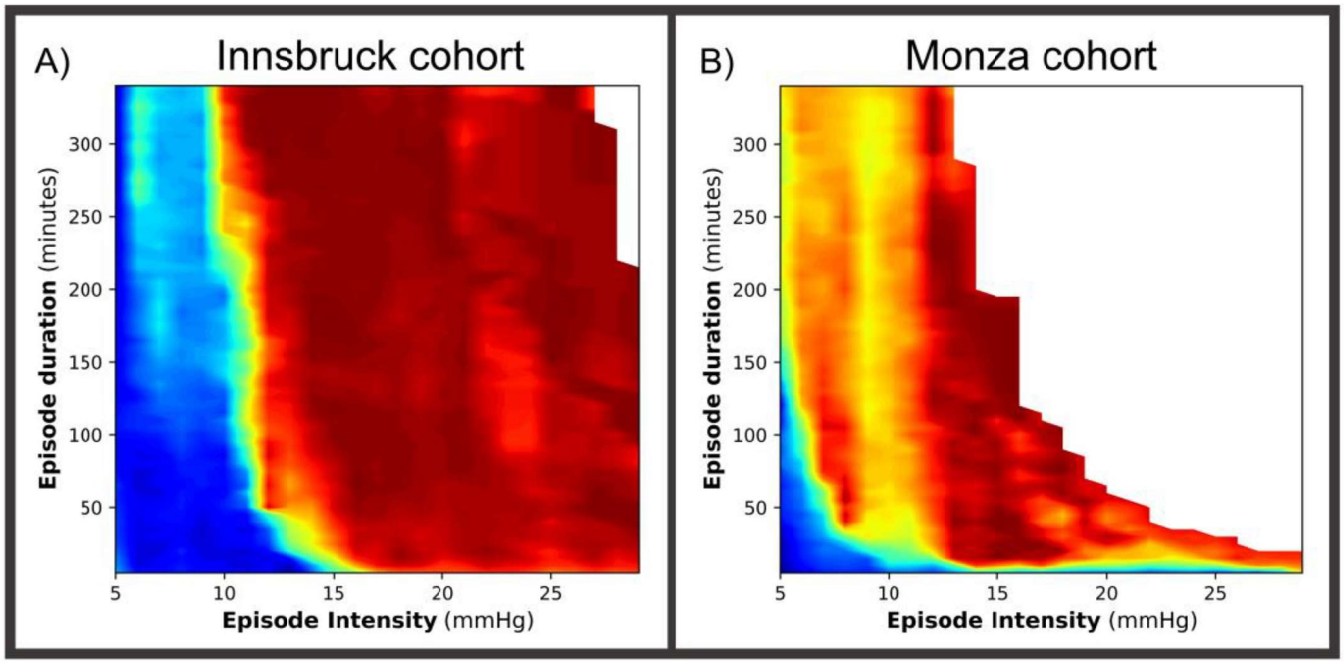


Figure 4.3 Visualization of the correlation between GOS and average number of ICP episodes of a certain dose, per GOS category. The episodes of dose (duration and intensity) of ICP that occur more frequently in patients with worse outcomes are represented in red. On the contrary, dark blue episodes mean that such episodes of dose of ICP, on average, occur more frequently in patients with better outcomes. Panel A shows the visualization obtained with the dataset from Innsbruck. Panel B shows the visualization obtained with the dataset from Monza.

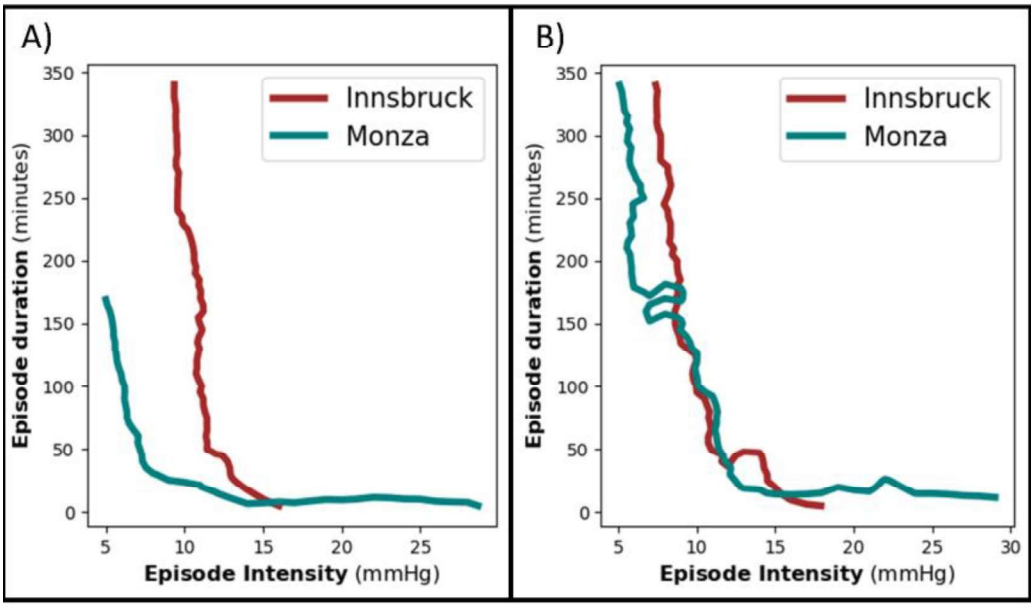


Figure 4.4 Comparison of transition curves of the two cohorts. The points of zero correlation between ICP doses and GOS compose the transition curve; hence, the transition curve indicates the line of transition between blue regions and red regions of the color-coded visualization. Panel A) transition curves obtained with the entire cohorts. The cohorts included 46 patients and 52 patients for Innsbruck and Monza, respectively. Panel B) transition curves obtained from subsets of patients that did not have rebleeding or delayed cerebral ischemia. The subsets included 26 patients and 28 patients for Innsbruck and Monza, respectively.

4.3.2 *Multivariable logistic regression analysis*

For both cohorts, the multivariable analysis showed an independent association between the ICP pressure-time burden and the 12-months neurological outcomes, *OR* -4.80, 95% *CI* -8.26 to -1.32, $p = 0.007$ and *OR* -5.11, 95% *CI* -8.38 to -1.84, $p = 0.002$ for Innsbruck and Monza, respectively, see Table 4.2.

Given that the average duration of monitoring time was different for the two centres, an additional analysis was performed for the Innsbruck cohort by limiting the recording to the first monitoring week. The independent association between the ICP pressure-time burden and neurological outcomes remained, $p = 0.03$, see Appendix 4.A.1 (Table 4.A.1).

It is noteworthy that no significant association was found between the percentage of monitoring time with ICP above 20 mmHg and neurological outcomes for either the Innsbruck (*OR* -0.03, 95% *CI* -0.07 to 0.01, *p* = 0.12) or Monza cohorts (*OR* -0.07, 95% *CI* -0.18 to 0.03, *p* = 0.18), see Table 4.3.

Table 4.2. Multivariable logistic regression analysis with covariate ICP pressure-time burden

INNSBRUCK			
Variable	OR	CI	*p*
Age	-0.06	[-0.14 to 0.01]	0.07
DCI	-1.09	[-2.68 to 0.50]	0.18
GCS	0.15	[-0.03 to 0.33]	0.11
ICP pressure-time burden	-4.80	[-8.26 to -1.32]	0.007
MONZA			
Variable	OR	CI	*p*
Age	-0.04	[-0.12 to 0.03]	0.28
DCI	-0.08	[-1.58 to 1.41]	0.91
GCS	0.04	[-0.12 to 0.20]	0.62
ICP pressure-time burden	-5.11	[-8.38 to -1.84]	0.002

DCI: delayed cerebral ischemia ; GCS: Glasgow Coma Scale; ICP: intracranial pressure

Table 4.3. Multivariable logistic regression analysis with covariate percentage of time with ICP > 20 mmHg

INNSBRUCK			
Variable	OR	CI	*p*
Age	-0.03	[-0.09 to 0.02]	0.23
DCI	-0.45	[-1.77 to 0.88]	0.51
GCS	0.13	[-0.03 to 0.30]	0.11
Percentage of time with ICP >20mmHg	-0.03	[-0.07 to 0.01]	0.12
MONZA			
Variable	OR	CI	*p*
Age	-0.04	[-0.11 to 0.02]	0.19
DCI	-0.80	[-2.17 to 0.56]	0.25
GCS	0.01	[-0.13 to 0.16]	0.85
Percentage of time with ICP >20mmHg	-0.07	[-0.18 to 0.03]	0.18

DCI: delayed cerebral ischemia ; GCS: Glasgow Coma Scale; ICP: intracranial pressure

4.4 DISCUSSION

In patients with aSAH, intracranial hypertension, defined as sustained ICP above 20-22 mmHg, is associated with increased mortality and worse neurological outcomes [4, 5, 21, 22]. It remains unclear whether the dose of ICP, which combines intensity and duration of an event of elevated ICP, is similarly associated with worse outcomes. In this study, we used the methodology introduced by Güiza et al [10] to investigate the association between ICP dose and long-term neurological outcomes of patients with aSAH. We found that the combination of intensity and duration defined the

tolerance to intracranial hypertension, suggesting that the concept of dose of ICP is valid in patients with aSAH.

In the visualizations, a semi-exponential transition clearly divided episodes of ICP associated with good outcomes from episodes associated with poor outcomes. Consistently with what has been reported in previous studies [4, 21, 22], Figure 4.3 shows the strong vulnerability of patients to intracranial hypertension. Such vulnerability can be partially explained by the poor clinical situation at admission, impaired cerebrovascular autoregulation [23] and the elevated age of the patients.

This study suggests that, in patients with aSAH, the association between intracranial hypertension and poor long-term neurological outcomes occurs at lower ICP values than what is hypothesized in the guidelines. This association towards lower ICP values can be partially explained by the continuously open EVD, which automatically results in lower ICP.

Alternatively, the results may indicate that intracranial hypertension is an important surrogate marker of underlying mechanisms, such as global brain edema, ischemia, cerebral infarction, hydrocephalus or brain metabolic crisis, that eventually lead to neurological deterioration [4].

The differences between the curves of the two cohorts could not be explained by baseline characteristics, but they might result from unaccounted effects of secondary complications such as re-bleeding, DCI, CSF dynamics disturbances, hematoma expansion or brain edema, which better define outcomes in patients with aSAH [21, 22]. Two supporting facts are the different incidence of rebleeding across the two cohorts and the increased similarity between the plots of the two centers when excluding patients with known secondary complications, see Figure 4.4. Patient management may also play a role in the definition of outcomes and hence explain the differences in the curves. In this study, minor differences were identified in the use of sedatives, vasopressors and amount of daily drained CSF. In addition, a small effect size was shown between the most frequent surgical procedure of the two centers. The confounding effect of secondary complications and treatment strategy is expected, given the complexity and heterogeneity of aSAH. In addition, the separate analysis of the two cohorts was crucial to identify this confounding effect, which might have been more pronounced in the Monza cohort, where the incidence of rebleeding is higher. A limitation of the visualization methodology is that it does not

allow corrections for confounders. Therefore, multivariable logistic regression analyses were used.

The main finding of this study is the independent association between the ICP pressure-time burden and the long-term neurological outcomes in aSAH patients. The ICP pressure-time burden was more predictive of long-term neurological outcomes than the time spent by the patient above the fixed threshold of 20 mmHg. The results, in agreement with previous studies on TBI [8, 10], suggest that the dose better quantifies the burden of intracranial hypertension than the time spent by the patient with an ICP above 20 mmHg.

This study has some limitations. First, it was performed on two cohorts of only 46 and 52 patients each, hence the results need to be validated in a bigger multicenter dataset. The use of a big multicenter database will be crucial to identify a universal ICP visualization curve on which draw general clinical conclusions. Second, this study is based on a retrospective analysis of retrospectively collected data of treated patients. As a consequence, the results show associations but not causal relations, and it is unclear whether intracranial hypertension causes poor neurocognitive outcomes, poor neurocognitive outcomes result from the occurrence of other secondary insults or the combination of both. Third, detailed information on the clinical management of the patients and on the occurrence of clinical complications (infections, seizures etc.) was unavailable. In this regard, the extraction of a universal ICP visualization curve will likely require the collection of a large and extensive dataset that includes all the information before mentioned. Last, the datasets collect continuous ICP and MABP recordings at different sampling frequencies and significantly different duration. Despite these limitations, multicenter studies with continuous ICP measurements of patients with aSAH, such as the one presented here, are exceptionally rare and therefore sorely needed [21].

The results of this study plead against the concept of a universal and fixed ICP threshold and suggest that the burden of elevated ICP might be better quantified with a more dynamic metric, such as the dose. Nonetheless, the time-intensity combinations on the transition should certainly not be viewed as treatment thresholds, as they represent an epidemiological finding in patients in whom the ICP was actively treated and of whom no detailed information on postoperative complications is available. Clinically relevant ICP dose intensity and duration thresholds still needs to be

defined, and are likely to be highly dependent on the presence of secondary complications and on the clinical strategy used for the management of aSAH.

4.5 CONCLUSION

The concept of association between ICP doses and neurological outcome, which has already been demonstrated in patients with TBI, may be valid in patients with aSAH. We have demonstrated that the burden of ICP dose is a strong predictor of neurological outcomes for patients with aSAH. Additional studies are needed to validate these findings.

ACKNOWLEDGMENT AND PERSONAL CONTRIBUTION

Study concept and design	Meyfroidt, Helbok, Citerio, Güiza, Depreitere
Data acquisition	Elli, Ianosi, Huber, Rass
Statistical analysis	**Carra,** Flechet
Interpretation of results	**Carra,** Meyfroidt, Güiza, Depreitere, Helbok, Citerio
Drafting of the manuscript	**Carra**
Manuscript revision	**All authors**
Principal investigator	Meyfroidt

4.A APPENDIX

4.A.1 Supplementary figures and tables

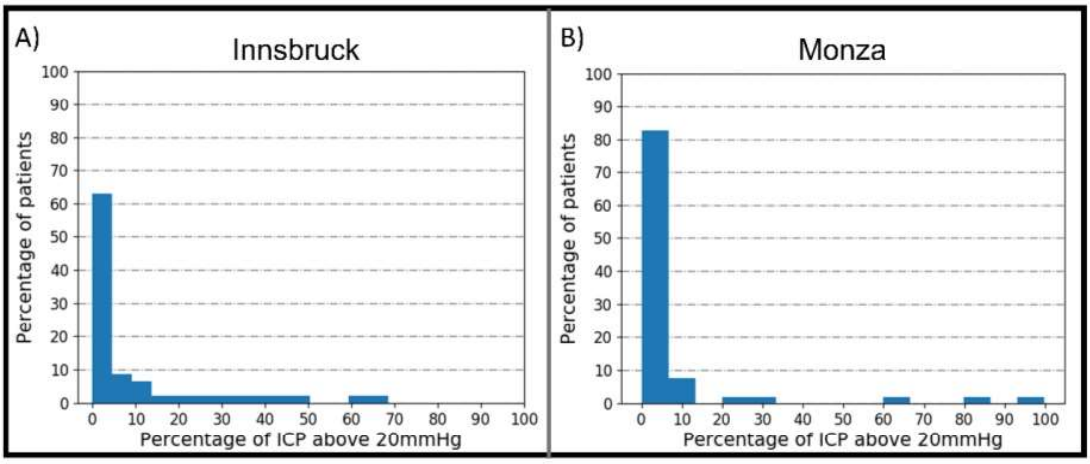


Figure 4.A.1 Percentage of monitoring time spent by the patient with an ICP above 20mmHg. Panel A) Innsbruck cohort. Panel B) Monza cohort. The number of patients is reported in percentages

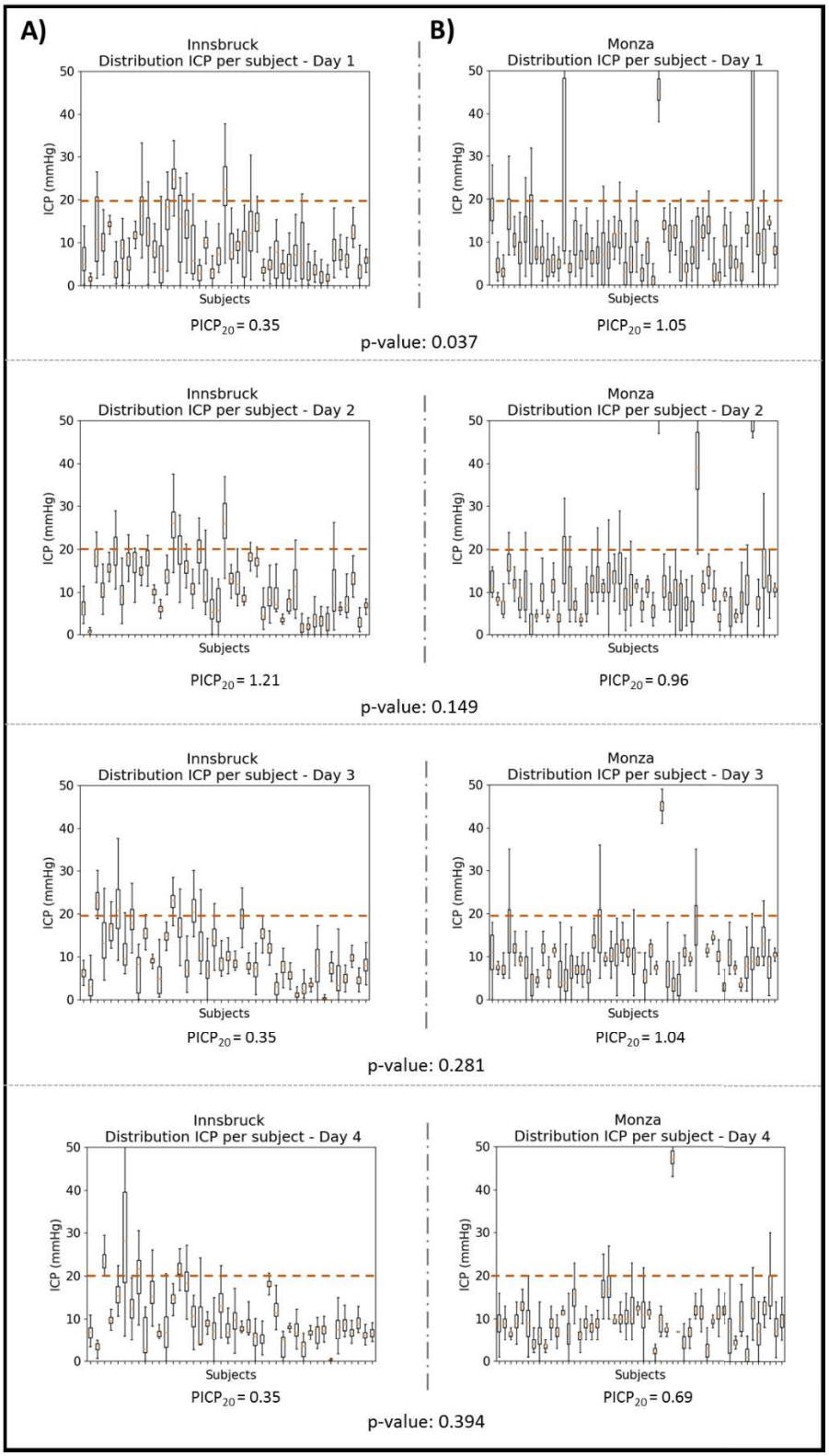


Figure 4.A.2 (part I) Distribution of the intracranial pressure (ICP) values per subject per day.

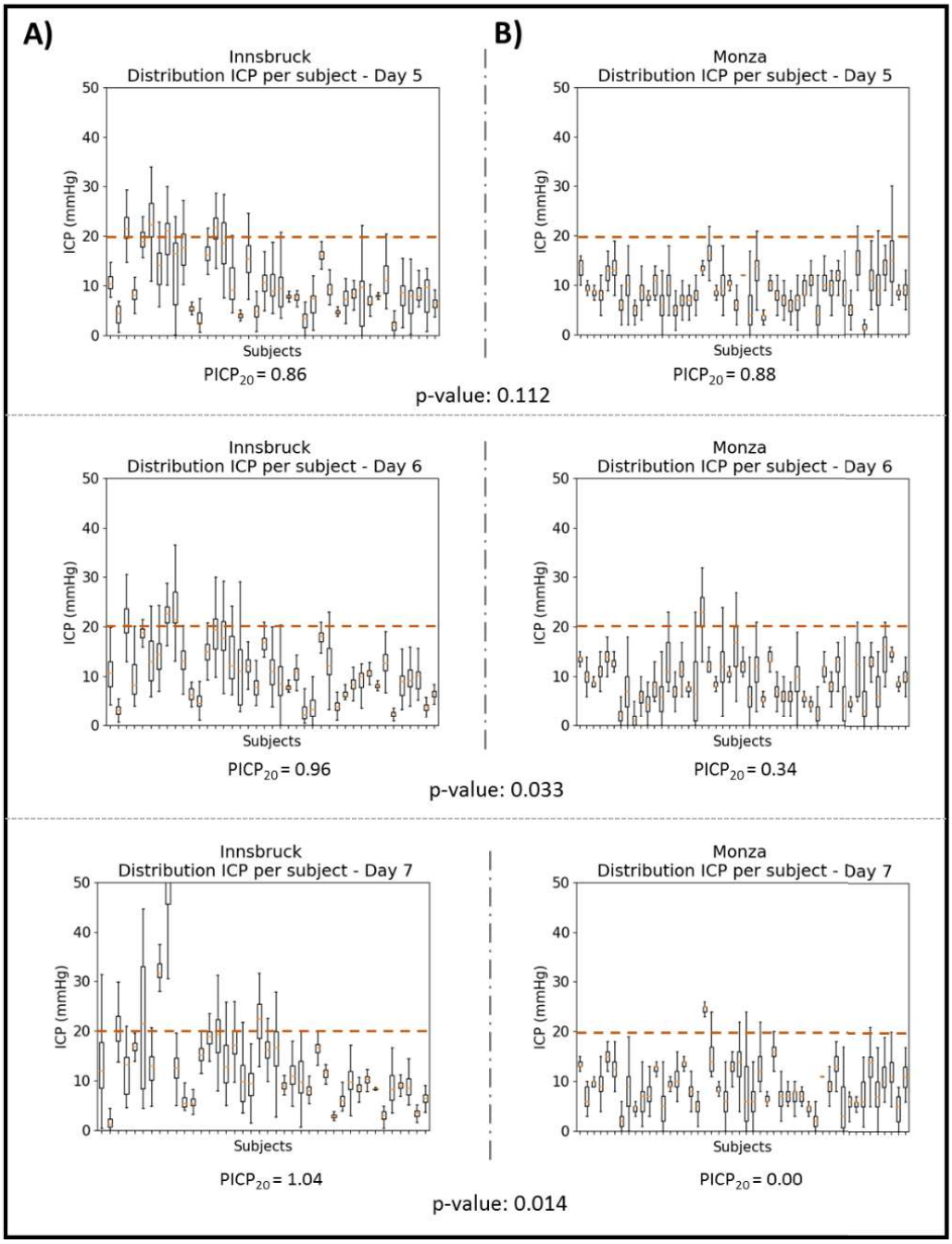


Figure 4.A.2 (part II) Distribution of the intracranial pressure (ICP) values per subject per day. Every subject is indicated with a boxplot, the box extends from the 25th to the 75th quartile, and the lower and top extremes of the whiskers represent the maximum and minimum values of the distribution. Panel A: Distribution of ICP values of the patients from Innsbruck cohort. Panel B: Distribution of ICP values of the patients from Monza cohort. The $PICP_{20}$ represents the median across the cohort of the percentage of monitoring time every patient spent above the threshold of 20 mmHg. The p-values of the Mann-Whitney U-test are reported.

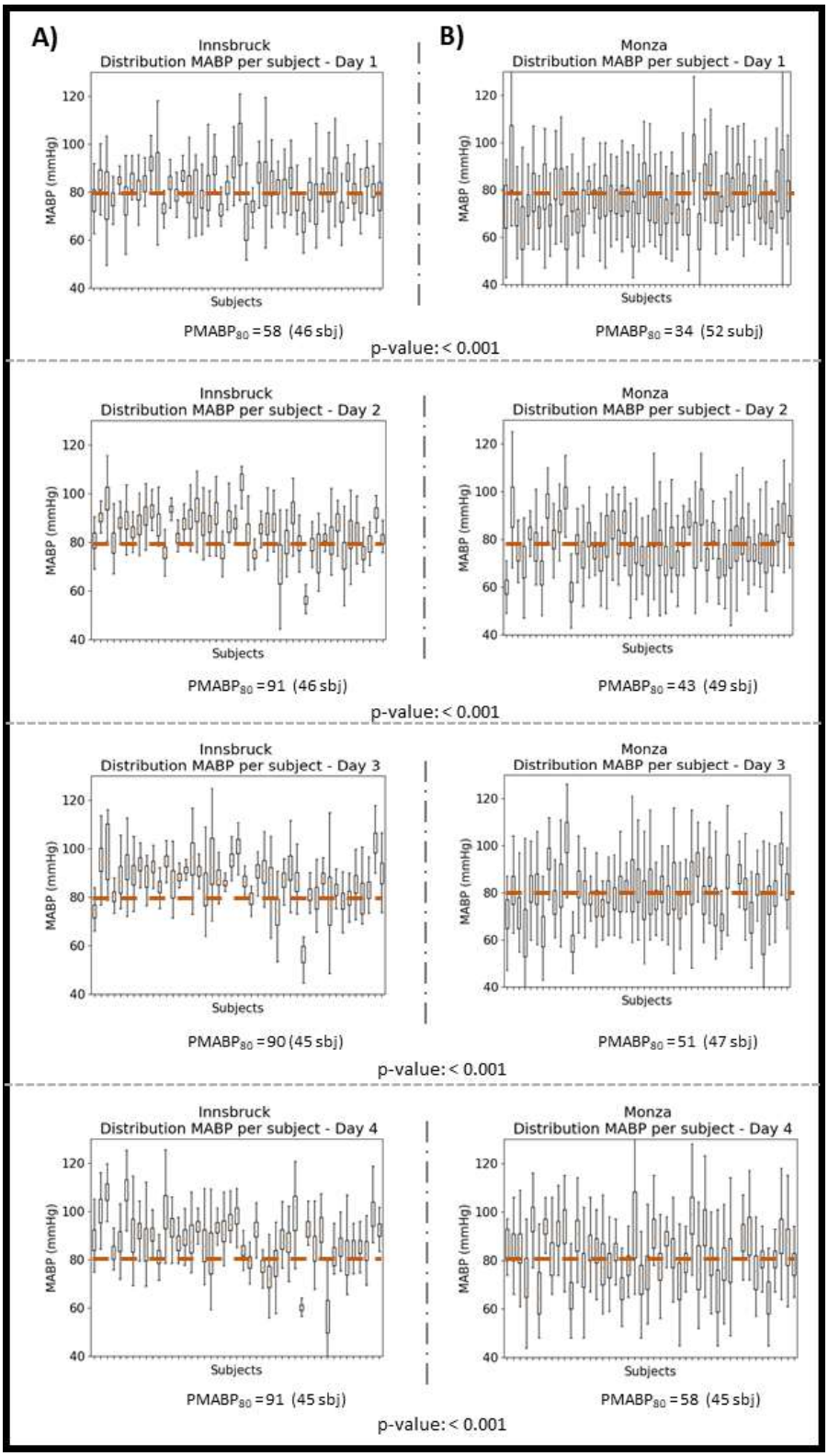


Figure 4.A.3 (part I) Distribution of the mean arterial blood pressure values (MABP) per subject per day.

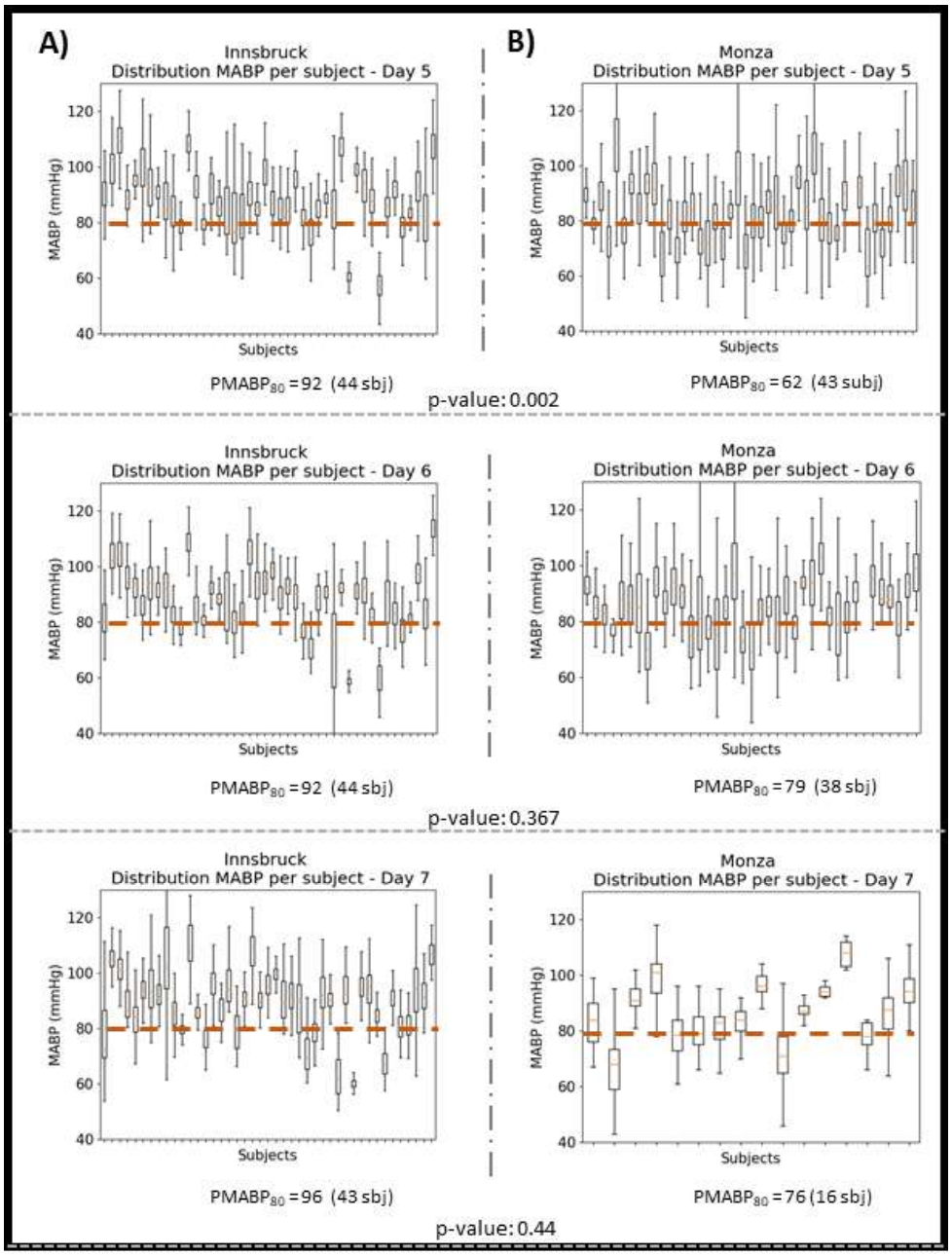


Figure 4.A.3 (part II) Distribution of the mean arterial blood pressure values (MABP) per subject per day. Every subject is indicated with a boxplot, the box extends from the 25th to the 75th quartile, and the lower and top extremes of the whiskers represent the maximum and minimum values of the distribution. Panel A: Distribution of MABP values of the patients from Innsbruck cohort. Panel B: Distribution of MABP values of the patients from Monza cohort. The $PMABP_{80}$ represents the median across the cohort of the percentage of monitoring time every patient spent above the threshold of 80 mmHg. The p-values of the Mann-Whitney U-test are reported.

Table 4.A.1 Multivariable logistic regression analysis, Innsbruck cohort, first week of monitoring

VARIABLE	OR	CI	*p*
Age	-0.06	[-0.13 to 0.004]	0.07
DCI	-1.01	[-2.48 to 0.46]	0.18
GCS	0.12	[-0.04 to 0.29]	0.15
ICP pressure-time burden	-3.59	[-6.85 to -0.32]	0.03

CI: confidence interval

DCI: delayed cerebral ischemia

4.A.2 *Analysis of the clinical management of the patients with aSAH for the different cohorts*

Differences in the clinical protocol for aSAH management were investigated. We analysed: the surgical strategy of each center for securing the ruptured aneurysm, the percentage of patients that underwent coiling and clipping (Figure 4.A.4), the percentage of patients that underwent vasopressors and sedatives daily (Figure 4.A.5, 4.A.6), the average amount of cerebrospinal fluid (CSF) that was drained daily from the patients (Figure 4.A.7, 4.A.8). The level of the EVD was not standardized between the two centers. Both centres use individual drainage levels, but in the Innsbruck center the drainage level usually starts at +5 cm, while in the Monza center the drainage level usually starts at +10 cm. These four variables were selected across the more complex range of factors that are part of aSAH management as the most important one to characterize aSAH management and to influence ICP values.

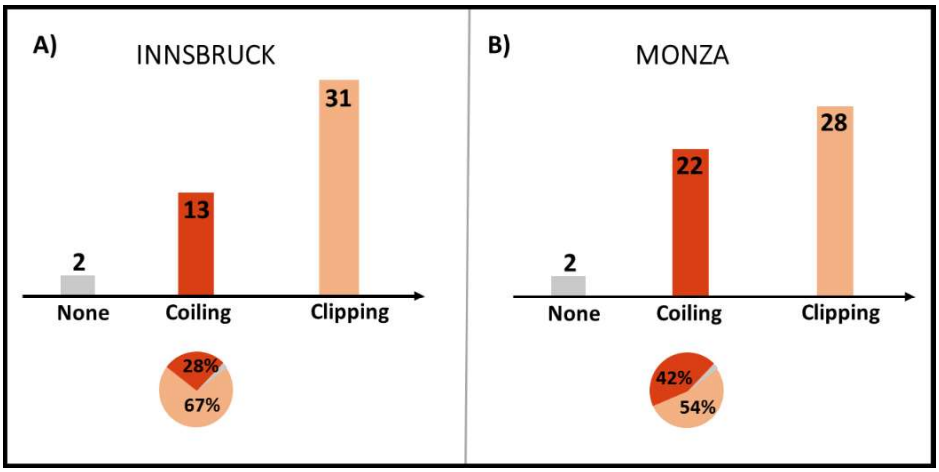


Figure 4.A.4 Distribution of patients that underwent coiling or clipping in the two populations. The proportions of coiling and clipping are not statistically significant different, $p = 0.13$ and $p = 0.17$ respectively. Panel A) Innsbruck cohort Panel B) Monza cohort

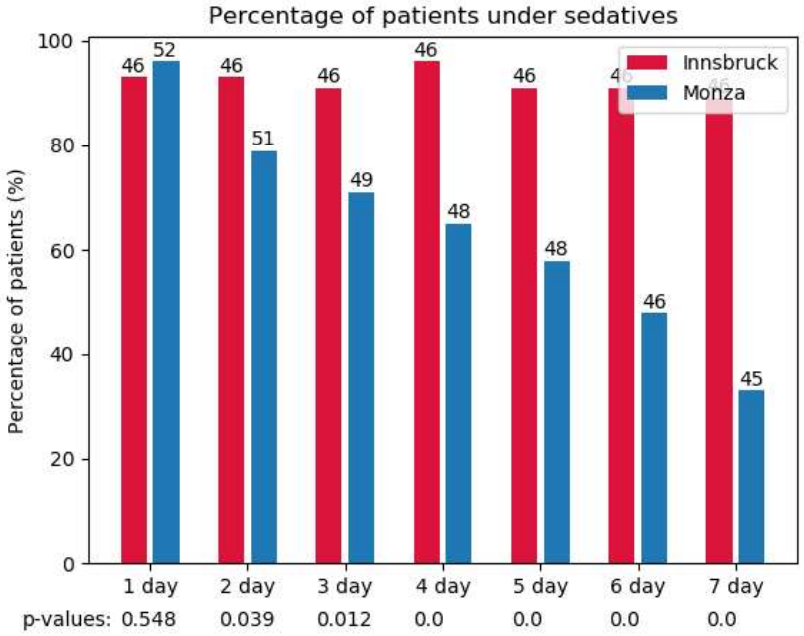


Figure 4.A.5 Percentage of patients that underwent sedatives daily. The analysis is limited to the first week of monitoring time. For every day of monitoring, the percentage of patients that underwent sedatives is indicated with a red bar for Innsbruck patients and with a blue bar for Monza patients. The percentage is computed by considering the total number of patients admitted in the study, the number of patients whose ICP was monitored that day is indicated on top of every bar.

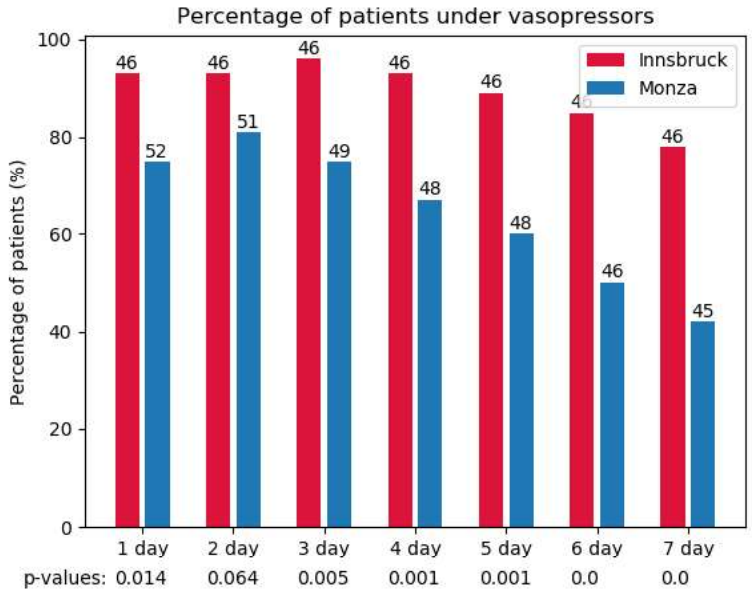


Figure 4.A.6 Percentage of patients that underwent vasopressors daily. The analysis is limited to the first week of monitoring time. For every day of monitoring, the percentage of patients that underwent vasopressors is indicated with a red bar for Innsbruck patients and with a blue bar for Monza patients. The percentage is computed by considering the total number of patients admitted in the study, the number of patients whose ICP was monitored that day is indicated on top of every bar.

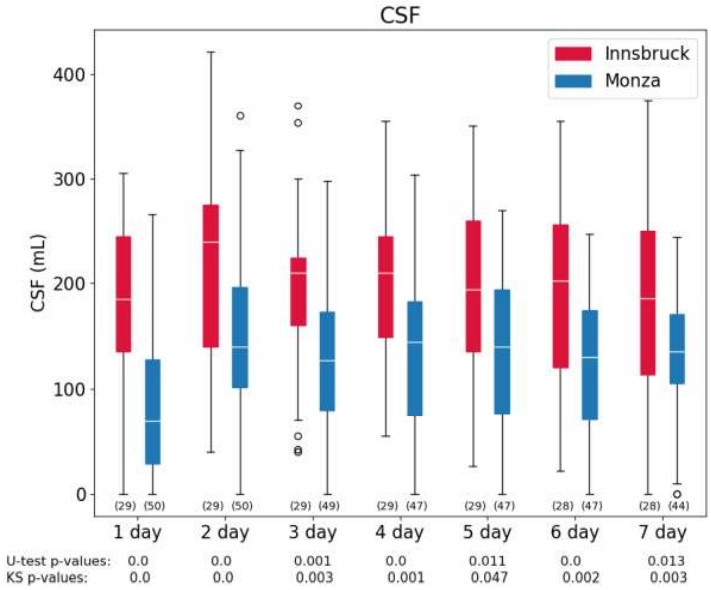


Figure 4.A.7 Comparison of the average CSF that was drained daily in the patients of the two cohorts. The analysis is limited to the first week of monitoring time. For

every day of monitoring, the distribution of the average CSF that was drained daily from every patient is indicated with a red boxplot for Innsbruck patients and with a blue boxplot for Monza patients. Below every boxplot it is reported the number of subjects with a valid CSF daily recording, i.e. the number of subjects that were used to extract the boxplots. The Mann-Whitney U-test and the Kolmogorov-Smirnoff (KS) test are reported at the bottom of the image to compare the two cohorts for every day of the first monitoring week.

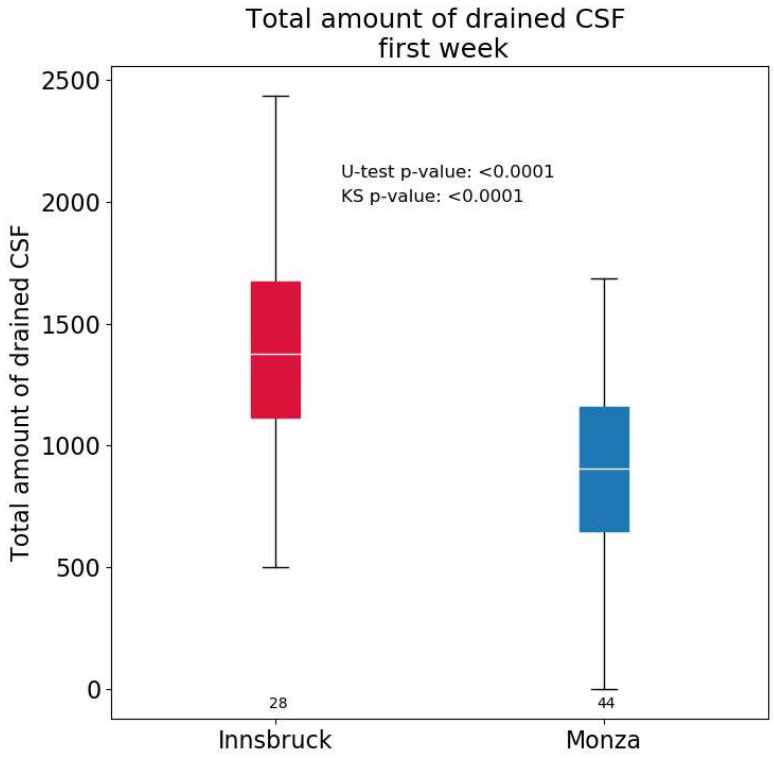


Figure 4.A.8 Boxplots of the total amount of cerebral spinal fluid that was drained from the patients during the first monitoring week. The cohorts are compared with the Mann-Whitney U-test and the Kolmogorov-Smirnoff (KS) test, whose values are reported in the figure and respectively equal to p <0.001 and p <0.001. Below every boxplot it is reported the number of subjects used to extract the boxplot.

BIBLIOGRAPHY

1. Etminan N, Chang H-S, Hackenberg K, et al (2019) Worldwide Incidence of Aneurysmal Subarachnoid Hemorrhage According to Region, Time Period, Blood Pressure, and Smoking Prevalence in the Population. JAMA Neurol. https://doi.org/10.1001/jamaneurol.2019.0006

2. Suarez J, Tarr R, Selman W (2006) Aneurysmal subarachnoid hemorrhage. N Engl J Med 354:387–396. https://doi.org/10.1056/NEJMra052732

3. Macdonald RL, Diringer MN, Citerio G (2014) Understanding the disease: aneurysmal subarachnoid hemorrhage. Intensive Care Med 40:1940–1943. https://doi.org/10.1007/s00134-014-3483-5

4. Zoerle T, Lombardo A, Colombo A, et al (2015) Intracranial pressure after subarachnoid hemorrhage. Crit Care Med 43:168–176. https://doi.org/10.1097/CCM.0000000000000670

5. Magni F, Pozzi M, Rota M, et al (2015) High-Resolution Intracranial Pressure Burden and Outcome in Subarachnoid Hemorrhage. Stroke 46:2464–2469. https://doi.org/10.1161/STROKEAHA.115.010219

6. (2007) Guidelines for the management of severe traumatic brain injury (3rd edition). J Neurotrauma 24:

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

8. Vik A, Nag T, Fredriksli O, et al (2008) Relationship of " dose " of intracranial hypertension to outcome in severe traumatic brain injury. J Neurosurg J Neurosurg J Neurosurg 109:0–678. https://doi.org/10.3171/JNS/2008/109/10/0678

9. Sheth KN, Stein DM, Aarabi B, et al (2013) Intracranial Pressure Dose and Outcome in Traumatic Brain Injury. Neurocrit Care 18:26–32. https://doi.org/10.1007/s12028-012-9780-3

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

11. Bederson JB, Connolly ES, Batjer HH, et al (2009) Guidelines for the Management of Aneurysmal Subarachnoid Hemorrhage: A Statement for Healthcare Professionals From a Special Writing Group of the Stroke Council, American Heart Association. Stroke 40:994–1025. https://doi.org/10.1161/STROKEAHA.108.191395

12. Helbok R, Olson DM, Le Roux PD, Vespa P (2014) Intracranial Pressure and Cerebral Perfusion Pressure Monitoring in Non-TBI Patients: Special Considerations. Neurocrit Care 21:85–94. https://doi.org/10.1007/s12028-014-0040-6

13. Frontera JA, Claassen J, Schmidt JM, et al (2006) Prediction of Symptomatic Vasospasmafter Subarachnoid Hemorrhage: The Modified Fisher Scale. Neurosurgery 59:21–27. https://doi.org/10.1227/01.NEU.0000218821.34014.1B

14. Hunt WE, Hess RM (1968) Surgical Risk as Related to Time of Intervention in the Repair of Intracranial Aneurysms. J Neurosurg 28:14–20. https://doi.org/10.3171/jns.1968.28.1.0014

15. (1988) Report of World Federation of Neurological Surgeons Committee on a Universal Subarachnoid Hemorrhage Grading Scale. J Neurosurg 68:. https://doi.org/10.3171/jns.1988.68.6.0985

16. Teasdale G, Jennett B (1974) Assessment of coma and impaired consciousness. A practical scale. Lancet (London, England) 2:81–4

17. Jennett B, Bond M, B.Jennett; M.Bond (1975) Assessment of Outcome After Severe Brain Damage. Lancet 1:480–484

18. Bederson JB, Connolly ES, Batjer HH, et al (2009) Guidelines for the management of aneurysmal subarachnoid hemorrhage: A statement for healthcare professionals from a special writing group of the stroke council, American heart association. Stroke 40:994–1025. https://doi.org/10.1161/STROKEAHA.108.191395

19. Diringer MN, Bleck TP, Claude Hemphill J, et al (2011) Critical Care Management of Patients Following Aneurysmal Subarachnoid Hemorrhage: Recommendations from the Neurocritical Care Society's Multidisciplinary Consensus Conference. Neurocrit Care 15:211. https://doi.org/10.1007/s12028-011-9605-9

20. Vargha A, Delaney HD, Vargha A (2000) A Critique and Improvement of the "CL" Common Language Effect Size Statistics of McGraw and Wong. J Educ Behav Stat 25:101. https://doi.org/10.2307/1165329

21. Cossu G, Messerer M, Stocchetti N, et al (2016) Intracranial pressure and outcome in critically ill patients with aneurysmal subarachnoid hemorrhage: A systematic review. Minerva Anestesiol 82:684–696

22. Heuer GG, Smith MJ, Elliott JP, et al (2004) Relationship between intracranial pressure and other clinical variables in patients with aneurysmal subarachnoid hemorrhage. J Neurosurg 101:408–416. https://doi.org/10.3171/jns.2004.101.3.0408

23. Budohoski KP, Czosnyka M, Kirkpatrick PJ, et al (2013) Clinical relevance of cerebral autoregulation following subarachnoid haemorrhage. Nat Rev Neurol 9:152–163. https://doi.org/10.1038/nrneurol.2013.11

5

PREDICTION MODEL FOR INTRACRANIAL HYPERTENSION DEMONSTRATES ROBUST PERFORMANCE DURING EXTERNAL VALIDATION ON THE CENTER-TBI DATASET

ADAPTED FROM: **Carra G.***, Güiza F.* et al. Prediction model for intracranial hypertension demonstrates robust performance during external validation on the CENTER-TBI dataset. Intensive Care Medicine, 2021 Jan; 47(1):124-126.

*contributed equally

Presented as: Poster presentation at the 33rd European Society of Intensive Care Medicine (ESICM) Annual Congress. Online, December, 2020.

Dear Editor,

In patients with traumatic brain injury (TBI), the management of elevated intracranial pressure (ICP) is mainly reactive, with ICP treated aggressively when it rises above 22 mmHg [1, 2].

To drive TBI care towards proactiveness, in 2013 Güiza et al. [3] developed a machine learning model for the prediction of extremely elevated ICP. Providing a 30 minutes forewarning, the model predicted with good discrimination and good calibration, events of ICP above 30 mmHg that lasted more than 10 minutes.

The model, which only requires continuous ICP and mean arterial blood pressure signals, was developed on data prior to 2005. Since changes in clinical practice can result in progressive degradation of model accuracy, the model was successively validated on a multicenter adult cohort collected between 2009 and 2011, with unchanged performance [4].

Similarly, the last years have seen further advances on TBI understanding and management, which led to the Seattle International Severe Traumatic Brain Injury Consensus Conference (SIBICC) in 2019 [1]. Hence, the need to validate the model once again, on new and independent data. In this study, we evaluated the performance of the model on the multicenter Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI) dataset [5].

The validation dataset included 257 patients with TBI, who were recruited prospectively between 2015 and 2017 as part of the CENTER-TBI high-resolution ICU monitoring cohort [5], see Table 5.1 for cohort demographics. Model performance was quantified with the metrics used in the previous publications [3][4], i.e. Area Under the Receiver Operating Curve (AUC), accuracy, sensitivity, specificity, and calibration analysis, while the clinical utility of the model was assessed through a decision curve analysis.

Table 5.1 Patients' demographic information for all cohorts

COHORTS	BRAIN-IT N=264	CENTER-TBI EXTERNAL VALIDATION N= 257	p value§
Age, years, median (IQR)	42 (26 to 58)	47 (30 to 61)	0.02
Sex, male, n (%)	212 (80)	208 (81)	0.87
Glasgow Coma Scale total, median (IQR)	7 (4 to 10)	6 (3 to 10)	0.14

§ p-values for continuous and ordinal variables with non-normal distribution (age, Glasgow Coma Scale) were computed with the Mann Whitney U-test, p-values for categorical variables (sex) were computed with the Chi-square test for proportions.

In this external validation dataset, which was collected almost 10 years later than the original development cohort, the model was still able to predict future episodes of extremely elevated ICP with good discrimination and calibration (AUC = 0.93, calibration slope 1.22, calibration–in–the–large = -0.04), see Figure 5.1. At the same cutoff of the original study, the model presented an accuracy of 88%, sensitivity of 83%, and specificity of 91%. These results demonstrate that the prediction model developed by Güiza et al. [3] is extremely robust to inter-center variability and most importantly, to progressive changes in clinical practice. The decision curve showed clinical usefulness for almost all risk thresholds (respectively from 4% to 96%), as shown in Figure 5.1. Acceptable alerting thresholds could range from 20% to 50% depending on the risk of the action taken in response to the alert, how the clinician values different outcomes for a specific patient, and the center-specific protocol for the management of elevated ICP. We believe that the model represents a promising and reliable tool for the proactive management of extremely elevated ICP, and that its early warnings on imminent ICP above 30 mmHg have the potential to improve clinical practice. Future prospective intervention studies are required to assess the impact of the use of this model at the bedside on mortality and patient outcome.

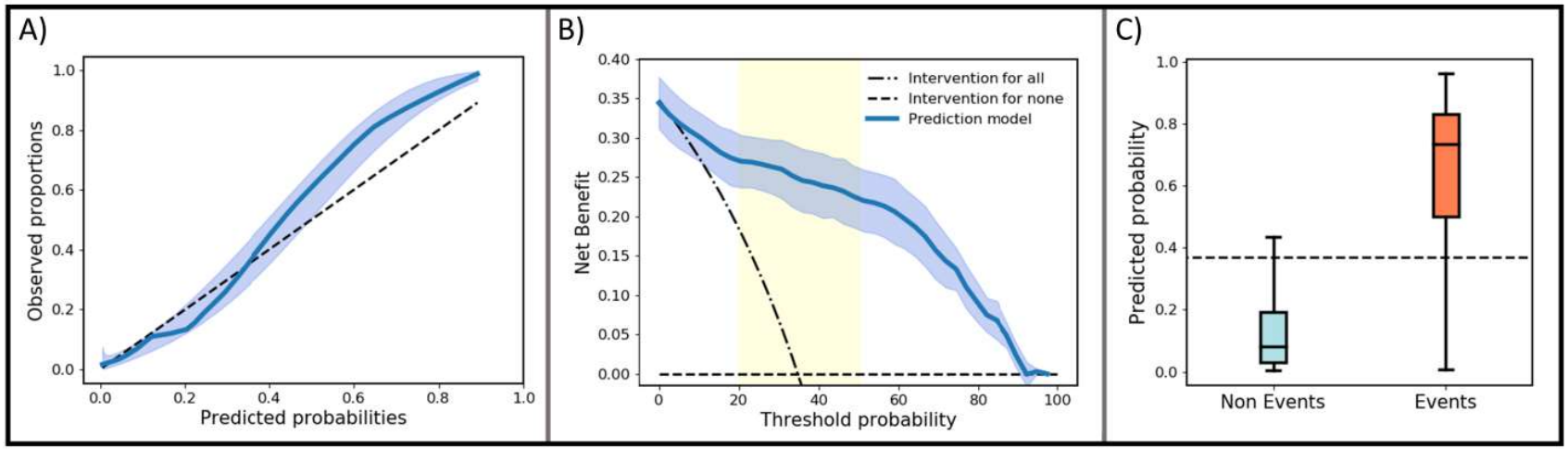


Figure 5.1 Performance metrics of the model developed by Güiza et al. for the prediction of elevated intracranial pressure on the CENTER-TBI dataset. Panel A) Calibration belt, represented with the shaded blue area, and calibration curve, represented with the blue line. Panel B) Decision curve of the prediction model, represented with the blue line, the shaded blue line represents the 95% confidence interval. The shaded yellow area indicates the range of possible alerting thresholds. Panel C) Boxplots of the predicted probabilities of the model for events of elevated ICP, in red, and events of not elevated ICP ("non events"), in cyan. The dashed black line indicates the probability cutoff used in the original study (equal to 0.37) to best discriminate between events and non-events.

ACKNOWLEDGMENT AND PERSONAL CONTRIBUTION

Study concept and design	Güiza, Meyfroidt, **Carra,** Depreitere
Data acquisition	CENTER-TBI high-resolution (HR-ICU) sub-study participants and investigators
Statistical analysis	**Carra**
Interpretation of results	**Carra,** Güiza, Meyfroidt
Drafting of the manuscript	**Carra**
Manuscript revision	**All authors**
Principal investigator	Meyfroidt

BIBLIOGRAPHY

1. Hawryluk GWJ, Aguilera S, Buki A, et al (2019) A management algorithm for patients with intracranial pressure monitoring: the Seattle International Severe Traumatic Brain Injury Consensus Conference (SIBICC). Intensive Care Med 45:1783–1794. https://doi.org/10.1007/s00134-019-05805-9

2. Smith M, Maas AIR (2019) An algorithm for patients with intracranial pressure monitoring: filling the gap between evidence and practice. Intensive Care Med 45:1819–1821. https://doi.org/10.1007/s00134-019-05818-4

3. Güiza F, Depreitere B, Piper I, et al (2013) Novel Methods to Predict Increased Intracranial Pressure During Intensive Care and Long-Term Neurologic Outcome After Traumatic Brain Injury: Development and Validation in a Multicenter Dataset. Crit Care Med 41:554–564. https://doi.org/DOI 10.1097/CCM.0b013e3182742d0a

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

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

6

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

ADAPTED FROM: **Carra G.** et al. "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". Submitted for publication.

ABSTRACT

BACKGROUND: 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. Nonetheless, the concept of dose is not used in clinical practice and not available as a bedside monitor metric, probably limited by its *a posteriori* calculation. Prediction of such harmful episodes of ICP dose could allow for a more proactive and preventive management of TBI. 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.

METHODS: 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 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.

FINDINGS: The external validation of the prediction model on the CENTER-TBI dataset demonstrated good discrimination and calibration (AUC: 0·94, accuracy: 0·88, sensitivity: 0·77, specificity: 0·93, calibration-in-the-large: 0·03, calibration slope: 0·93). Similar performance were obtained when the model was tested for different forewarning times.

INTERPRETATION: 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.

6.1 INTRODUCTION

Patients with severe Traumatic Brain Injury (TBI) may suffer from prolonged periods of elevated intracranial pressure (ICP),[1] with an increased risk of mortality and poor neurological outcomes [2–4].

Current guidelines suggest to treat the ICP when it rises above 22 mmHg [5], but such threshold-based clinical strategy presents some important limitations [6–9]. First, the proposed threshold is population-derived, therefore it does not allow therapy to be targeted to specific subgroups of patients. On this note, the cohort study [10] that led to the most recent change in threshold recommendations from 20 to 22 mmHg actually identified different thresholds for elderly and female patients. Second, some studies have shown that ICP levels lower than 22 mmHg might also associate with worse neurological outcomes [11, 12]. Third, secondary injury by elevated ICP is not adequately defined by the simple, sometimes brief, 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 secondary brain injury due to elevated ICP. High doses of elevated ICP have been associated with worse clinical outcomes in several observational studies [11, 13–15], indicating that changes in ICP over time may have a greater clinical relevance than time-point ICP values above a fixed threshold [14, 16].

The visualization of the association between different ICP intensities and durations, namely ICP doses, and long-term neurological outcomes, as described in a color coded heat map, was initially proposed by Güiza et al. [11], see Figure 6.1 panel C. The visualization has been further replicated in other large datasets [12, 17]. An exponential line separates the ICP doses that occur more frequently in patients with worse neurological outcomes (lower Glasgow Outcome Score [GOS] [18]) from the doses that occur more frequently in patients with better outcomes (higher GOS). These studies suggest that high ICP values can be tolerated if maintained for a short period. On the contrary, ICP values between 15 mmHg and 22 mmHg, if maintained for a prolonged time, could still be associated with poor neurological outcomes. Panel A) and B) of Figure 6.1 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.

In spite of this previous work [11, 14], the concept of ICP dose has not reached clinical practice. One limitation could be its retrospective calculation. As such, the dose of ICP informs the clinician on past ICP events (which are associated with possible patient outcomes), but has little effect on the planning of the future therapeutic strategy. In this context, early-warnings of impending harmful doses of ICP could be valuable information for the attending clinician [7]. Several prediction models have attempted to predict ICP elevations [19–23] but none of them are currently in use at the bedside.

Given the complexity of TBI pathophysiology, ICP alone cannot provide complete knowledge of the status of the brain [2, 24, 25]. For instance, the mean arterial blood pressure (MAP) is an important contributor to the perfusion status of the brain. Local brain oxygenation can be monitored with the partial pressure of brain tissue oxygen ($PbtO_2$). Similarly, cerebrovascular autoregulation (CAR) plays a crucial role in fostering oxygen and nutrient supply. CAR is a dynamic phenomenon that is difficult to assess in a clinical context [26] and for which computational indices, such as the Pressure Reactivity index (PRx) [27] or the low-frequency autoregulatory index (LAx) [28], have been proposed. These indices, which are computed from the ICP and MAP signals, are currently only available in selected research settings in the form of research tools (ICM+, Cambridge Enterprise Ltd., UK).

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

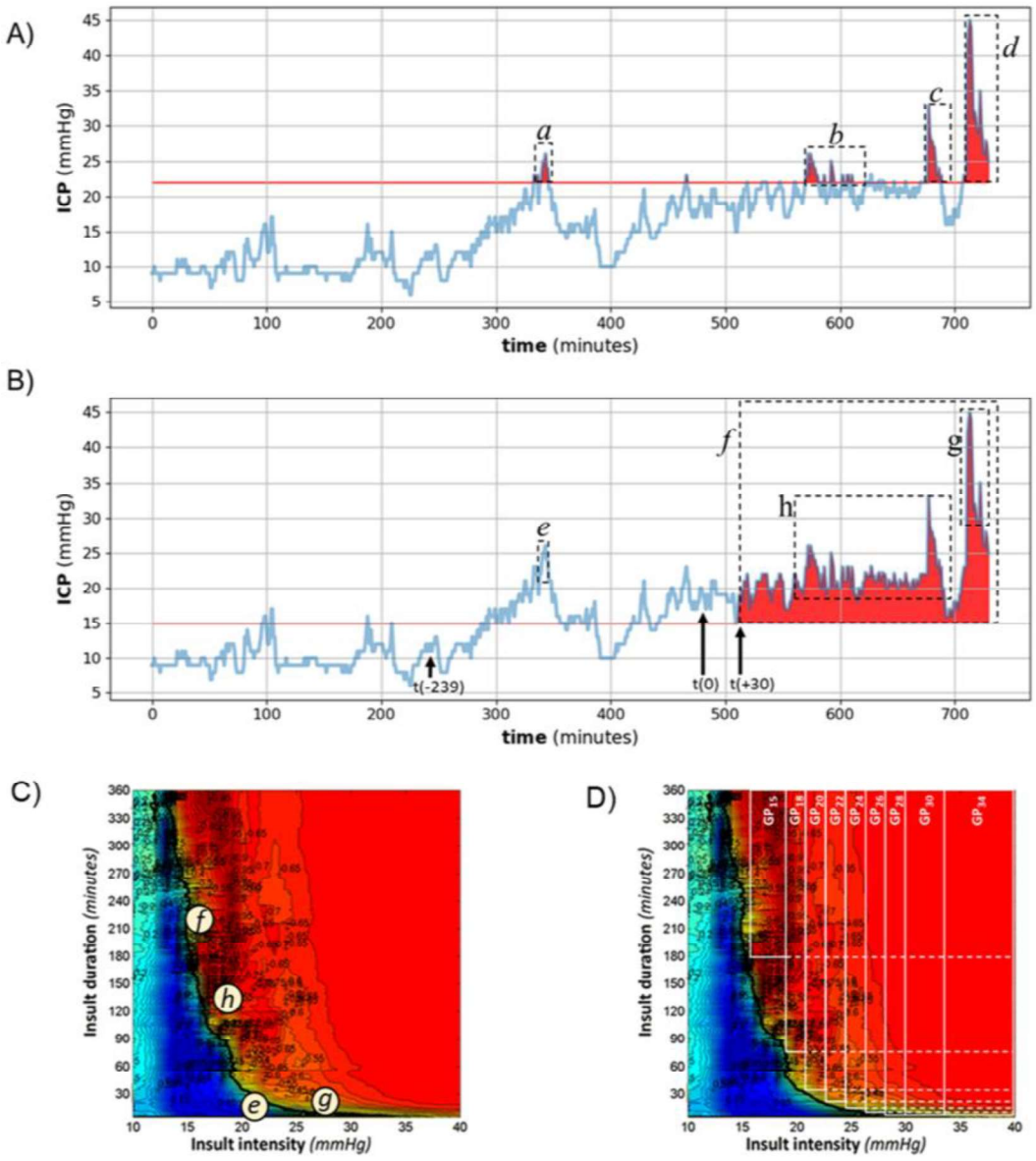


Figure 6.1 Panel A) and B) Visualization of ICP harmfulness as evaluated according to different criteria. Panel A) ICP harmfulness is defined by events of ICP>22 mmHg. Events of ICP that meet this criterion are identified with the letters *a*, *b*, *c* and *d*. Panel B) ICP harmfulness is defined by events of ICP in the "red area" of the visualization proposed by Güiza et al. [11] (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>15 mmHg that lasts more than 180 minutes, *g* is an event of ICP>28 mmHg that lasts more than 10 minutes, and *h* is an event of ICP>18 mmHg 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 22 mmHg but not long enough (less than 25 minutes) to be 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. [11]. 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). Panel D) visual representation of the division of the "red area" of the visualization of Güiza et al. [11] in sub-areas. Every sub-area extends until the 40 mmHg 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.

6.2 METHODS

6.2.1 *Database*

The development cohort included the data of 290 patients from 6 prospectively and retrospectively collected databases: Brain-IT [29] 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 were admitted in the San Gerardo Hospital in Monza, Italy between March 2010 and April 2013; 27 from the University Hospitals of Leuven, Belgium between September 2010 and September 2013; 19 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 [30]. 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 2 consecutive values were imputed with the median value of the previous 10 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 [31, 32] in use during data acquisition.

6.2.2 *Predictive task*

The model predicts with 30 minutes forewarning a broad set of events of ICP whose dose (combination of intensity and duration) was associated with poor long-term neurological outcomes in the visualization method proposed by Güiza et al. [11], as represented by the red area of Figure 6.1, panel C. Given that the predictive patterns that precede the different harmful ICP doses may differ, we divided the red area into several sub-areas and targeted each sub-area separately. The red sub-areas were defined 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 > 24 mmHg 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 6.1, panel D. The combination of these red sub-areas create the prediction target of the model, which is a simplified representation of the red area proposed by Güiza et al. [11] .

6.2.3 *Model development*

The development of the model for the prediction of ICP events in the entire red area was performed as follows:

1. First, a prediction model was developed for each specific red sub-area. To predict doses of ICP > 30 mmHg for more than 10 minutes, we used the prediction model previously proposed by

Güiza et al. [23]. This Gaussian Processes (GP) regression model was similarly developed on the Brain-IT [29] database and presented robust performance when externally validated on a large adult cohort and on the CENTER-TBI dataset [33, 34] .

2. Afterwards, we developed a model that combines the predictions of the models for the red sub-areas and provides an integrated prediction. This integrated prediction does not refer to a specific region of the red area, but it generally refers to ICP doses in the red area.

To increase generalizability and avoid overfitting, each model was trained with 10-fold cross validation (CV).

6.2.4 *Prediction models for red sub-areas*

The prediction models for the red sub-areas were developed with GP Regressor with a Rational Quadratic kernel function. The use of a GP Regressor was due to the good performance that this method obtained in previous studies [34]. 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 6.1 panel D.

Input features 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 6.1 panel B. For a complete list of the extracted features see Appendix 6.A.1.

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 [35] and features selection via mutual information [36, 37]. Feature importance was computed with the permutation importance technique [38].

6.2.5 *Prediction model for the entire red-area*

The model for the prediction of the broad set of ICP events in the red area is a Random Forest (RF) classifier. For this task, the RF classifier

demonstrated superior performance as compared to other tested models (linear regression model, GP classifier and decision tree classifier). For simplicity, we will further refer to this model as RFred model. The RFred model provides as output the probability that the patient will experience events of ICP in the red area in the next 30 minutes. We believe this output type, *i.e.* probability of being in the red area, is particularly congenial to the clinical environment.

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.

6.2.6 *External validation*

To assess generalizability, the GPx_s and RFred models were externally validated on the 264 patients from the CENTER-TBI dataset [30]. External validation is a crucial step in model development, given that 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.

We used the external validation dataset to perform an additional exploratory analysis. In detail, we assessed changes in performance of the model when predicting events of ICP in the red area with different forewarning times, namely 5 minutes, 15 minutes, 45 minutes and 60 minutes. This exploratory analysis aimed at providing insights on the performance of the model when used in real time, where the forewarning time of 30 minutes might not strictly apply and where deleterious episodes of ICP dose could be potentially predicted even earlier.

6.2.7 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) of 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/). Statistical analysis, model development and model evaluation were performed 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).

6.2.8 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 [39]. In addition, 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. [40].

6.3 RESULTS

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 of the GPx and RFred models. In short, 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 information on the performance of the GPx and RFred models on the development cohort and feature importance analysis can be found in Appendix 6.A.2, 6.A.3, 6.A.4 and 6.A.5.

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 6.1 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 0·89, an accuracy of 0·88, a precision of 0·85, a sensitivity of 0·77 and a specificity of 0·93. 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 6.2, 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 6.2, panel B. Table 6.2 summarizes the main characteristics of the model in the form of a model fact sheet.

When tested for different forewarning times, the model presented a classification accuracy of 0.89, 0.89, 0.88 and 0.88 for respectively a 5 minutes, 15 minutes, 45 minutes and 60 minutes forewarning time. For the complete list of the performance metrics, see Appendix 6.A.6.

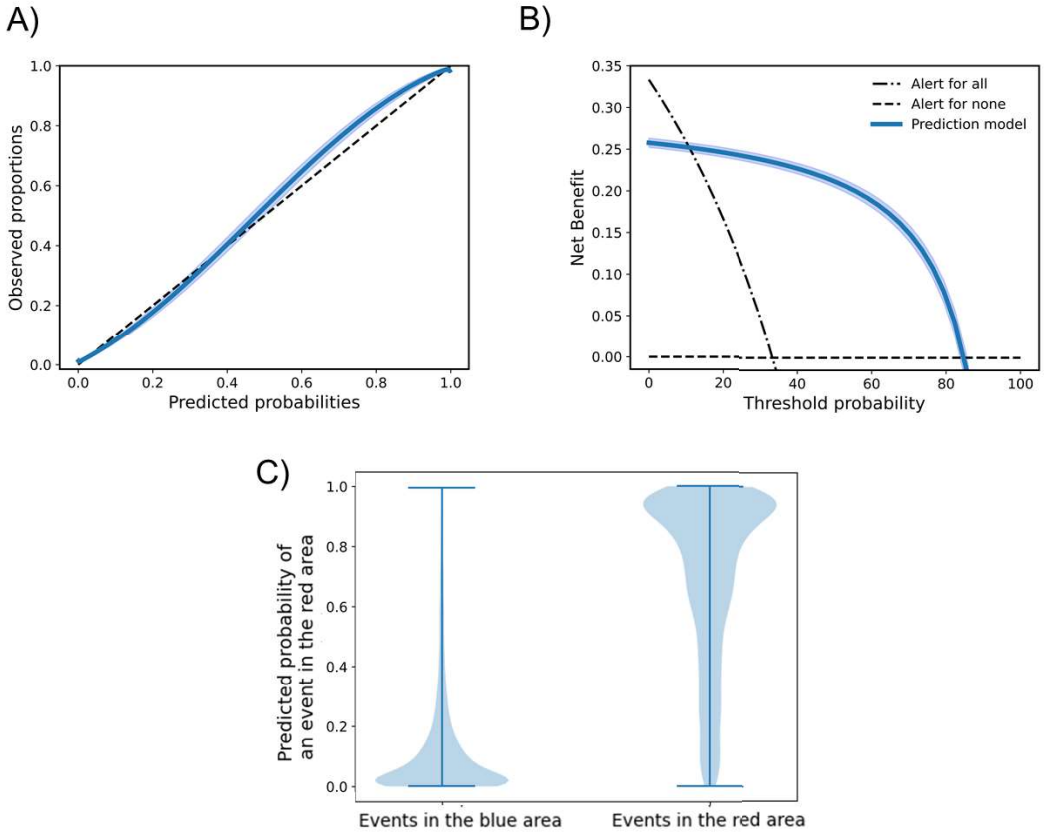


Figure 6.2 Performance metrics of the RFred model for the prediction of the red area on the external validation dataset. Panel A) calibration curve. Panel B) decision curve. Panel C) violin plots of the predicted probabilities for events in the red and blue area.

Table 6.1 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	0·89	0·88	0·85	0·77	0·93	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 computed 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 6.2 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 provides alerts for the early identification of future events of potentially harmful ICP doses and therefore it alerts the clinician when a patient is in need of specific medical attention.

6.4 DISCUSSION

In this study, we present a machine-learning model for the prediction 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 [19, 21, 22, 34], with common characteristics. Several of these models focused on one specific ICP insult of specific intensity and duration, not taking into consideration the complex, broad range of ICP events that have been associated with poor outcomes [19, 34]. Moreover, most of these prediction models use a short forewarning time, which may be insufficient to trigger a clinically useful intervention [21, 22]. An additional characteristic that could challenge a potential clinical implementation is the large number of required inputs, often from multiple monitoring sources, which not only obstacle the transfer to a clinical setting but also increases the risk of overfitting [22]. Last, but most importantly, most of these models lack external validation on geographically and temporally independent datasets [19, 21, 22]. External validation is strongly recommended [39, 41, 42], 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 0.89, an accuracy of 0.88, a precision of 0.85, a sensitivity of 0.77 and a specificity of 0.93. As this large, external, multicenter dataset was collected more than 10 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 over time. 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 intracranial hypertension [11]. This target was based on a previous study [11], but similar associations with outcome were observed in a large single-center cohort [12] and in the CENTER-TBI dataset [17]. 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.

Third, the forewarning time interval was defined after consultation with 3 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. Importantly, the exploratory analysis on the CENTER-TBI dataset showed that performance does not change remarkably for different forewarning time. This suggests that the model could provide prediction of harmful doses of ICP even earlier. In addition, it proves the robustness of the model. Future studies are needed to assess the role of the forewarning time when using the model in real time.

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 Güiza et al. [11] that describes the association between ICP doses and long term neurological outcomes on the general population. However, 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..) [11]. 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. Second, the data used were all acquired in European settings. Validation on datasets collected in non-European ICUs are therefore necessary to fully assess generalizability. Third, 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 [43]. 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. However, information on the CAR status of the patient was included as input to the models in the form of LAx signal. The fourth limitation of this study is linked to the lack of absolute explainability of machine learning models, which may limit the degree to which it is trusted and accepted by nurses and clinicians. To overcome this limitation, we performed a feature importance analysis for each GP 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. 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. 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 this 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 long-term neurocognitive outcomes.

6.5 CONCLUSIONS

In this study we present an accurate, sparse and robust model for the prediction 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 minute 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.

ACKNOWLEDGMENT AND PERSONAL CONTRIBUTION

Study concept and design	Güiza, Meyfroidt, **Carra,** Depreitere
Data acquisition	Piper on behalf of the BRAIN-IT collaboration, Citerio on behalf of the San Gerardo hospital of Monza (Italy), Maas on behalf of the NEMO project and the CENTER-TBI high-resolution (HR-ICU) sub-study participants and investigators
Statistical analysis	**Carra**
Interpretation of results	**Carra,** Güiza, Meyfroidt
Drafting of the manuscript	**Carra**
Manuscript revision	**All authors**
Principal investigator	Meyfroidt

6.A.1 Complete list of the extracted features

Table 6.A.1 lists the 228 potential predictors that were generated from the intracranial pressure (ICP), mean arterial blood pressure (MAP) and low frequency autoregulatory index (LAx)[44] signals. These included: a) median value and standard deviation of consecutive non-overlapping intervals of the continuous signal b) frequency-components analysis of the continuous signal c) presence of previous events of ICP doses in the red area, in the 4 hours preceding the prediction. Features were extracted if the amount of missing values did not exceed the 10% of the duration of the prediction window (of duration 4 hours).

The LAx is computed from the ICP and MAP continuous signals [44], for computational reasons LAx features are extracted only from the last 2h of monitoring preceding the prediction.

Table 6.A.1 List of features extracted from the ICP, MAP and LAx signals.

NUMBER OF FEATURES	FEATURE DESCRIPTION
24	Median of ICP across consecutives, non-overlapping windows of duration 10 minutes during the 4h window
24	Median of MAP across consecutives, non-overlapping windows of duration 10 minutes during the 4h window
12	Median of LAx across consecutives, non-overlapping windows of duration 10 minutes during the 2h window
24	Standard deviation of ICP across consecutives, non-overlapping windows of duration 10 minutes during the 4h window
24	Standard deviation of MAP across consecutives, non-overlapping windows of duration 10 minutes during the 4h window
12	Standard deviation of LAx across consecutives, non-overlapping windows of duration 10 minutes during the 2h window
8	Median of ICP across consecutives, non-overlapping windows of duration 30 minutes during the 4h window
8	Median of MAP across consecutives, non-overlapping windows of duration 30 minutes during the 4h window
4	Median of LAx across consecutives, non-overlapping windows of duration 30 minutes during the 2h window

8	Standard deviation of ICP across consecutives, non-overlapping windows of duration 30 minutes during the 4h window
8	Standard deviation of MAP across consecutives, non-overlapping windows of duration 30 minutes during the 4h window
4	Standard deviation of LAx across consecutives, non-overlapping windows of duration 30 minutes during the 2h window
4	Median of ICP across consecutives, non-overlapping windows of duration 60 minutes during the 4h window
4	Median of MAP across consecutives, non-overlapping windows of duration 60 minutes during the 4h window
2	Median of LAx across consecutives, non-overlapping windows of duration 60 minutes during the 2h window
4	Standard deviation of ICP across consecutives, non-overlapping windows of duration 60 minutes during the 4h window
4	Standard deviation of MAP across consecutives, non-overlapping windows of duration 60 minutes during the 4h window
2	Standard deviation of LAx across consecutives, non-overlapping windows of duration 60 minutes during the 2h window
5	First 5 coefficients of the cepstrum of the ICP signal during the 4h window
5	First 5 coefficients of the cepstrum of the MAP signal during the 4h window
5	First 5 coefficients of the cepstrum of the LAx signal during the 2h window
5	Largest 5 frequency components of the Fourier transform of the ICP signal during the 4h window
5	Largest 5 frequency components of the Fourier transform of the MAP signal during the 4h window
5	Largest 5 frequency components of the Fourier transform of the LAx signal during the 2h window
4	Frequency of the largest 5 frequency components of the Fourier transform of the ICP signal during the 4h window (excluding the 0Hz frequency)
4	Frequency of the largest 5 5 frequency components of the Fourier transform of the MAP signal during the 4h window (excluding the 0Hz frequency)
4	Frequency of the largest 5 frequency components of the Fourier transform of the LAx signal during the 2h window (excluding the 0Hz frequency)
1	Presence of episodes of ICP doses in the targeted red sub-area in the 4h window (yes/no variable)
1	Number of episodes of ICP doses in the targeted red sub-area in the 4h window

1	Average permanence in the targeted red sub-area in the 4h window
1	Presence of episodes of ICP doses of high intensity but too short duration to belong the targeted red sub-area in the 4h window (yes/no variable)
1	Number of episodes of episodes of ICP doses of high intensity but too short duration to belong the targeted red sub-area in the 4h window (yes/no variable)
1	Percentage of time with passive autoregulation in the 2h window

6.A.2 *Performance of the GPx models for the prediction of the red sub-areas*

Table 6.A.2 reports the number of events per patient per red sub-area category.

Table 6.A.2 Number of events in each identified red-sub-area

MODEL	NUMBER OF EVENTS IN THE DEVELOPMENT COHORT (TOTAL SIZE)	NUMBER OF EVENTS PER PATIENT MEDIAN [IQR]
GP_{34}	299 (897)	0 [0 to 1]
GP_{28}	827 (2481)	1 [0 to 3]
GP_{26}	715 (2145)	1 [0 to 3]
GP_{24}	703 (2109)	1 [0 to 3]
GP_{22}	633 (1899)	1 [0 to 3]
GP_{20}	801 (2403)	1 [0 to 3]
GP_{18}	759 (2277)	1 [0 to 2]
GP_{15}	678 (2034)	1 [0 to 1]

Summary of the performance of the prediction models for the red sub-area are presented in Table 6.A.3. On the 10 folds CV internal validation sub-sets, all GP models presented a mean AUC above 0·86, a mean AP above 0·64, a mean accuracy above 0·79, a mean precision above 0·64, a mean sensitivity above 0·81 and a mean specificity above 0·75. The GP models presented a mean calibration-in-the-large included between -0·05 and -0·01 and mean calibration slope included between 1·03 and 1·09. The calibration curves p-values were equal to 0·069, 0·000, 0·000, 0·000, 0·000, 0·000, 0·011, 0·000 for respectively GP_{34}, GP_{28}, GP_{26}, GP_{24}, GP_{22}, GP_{20}, GP_{18}, GP_{15}.

Table 6.A.3 Performance of the GPx models for the prediction of the red subareas

MODEL	AUC mean (SD)	AP mean (SD)	ACCURACY mean (SD)	PRECISION mean (SD)	SENSITIVITY mean (SD)	SPECIFICITY mean (SD)	CALIBRATION-IN-THE-LARGE mean (SD)	CALIBRATION SLOPE mean (SD)
GP_{34}	0·92 (0·02)	0·86 (0·03)	0·87 (0·02)	0·86 (0·03)	0·85 (0·04)	0·88 (0·01)	-0·03 (0·03)	1·08 (0·08)
GP_{28}	0·89 (0·01)	0·79 (0·03)	0·83 (0·02)	0·71 (0·05)	0·83 (0·03)	0·82 (0·05)	-0·03 (0.03)	1·08 (0·07)
GP_{26}	0·86 (0·03)	0·74 (0·05)	0·80 (0·02)	0·66 (0·04)	0·81 (0·06)	0·79 (0·04)	-0·01 (0·05)	1·03 (0·10)
GP_{24}	0·88 (0·02)	0·77 (0·04)	0·81 (0·02)	0·68 (0·04)	0·83 (0·05)	0·80 (0·03)	-0·02 (0·03)	1·05 (0·07)
GP_{22}	0·86 (0·02)	0·71 (0·04)	0·79 (0·04)	0·65 (0·06)	0·82 (0·06)	0·78 (0·07)	-0·01 (0·05)	1·06 (0·13)
GP_{20}	0·87 (0·01)	0·64 (0·04)	0·79 (0·02)	0·64 (0·04)	0·85 (0·03)	0·76 (0·04)	-0·05 (0·02)	1·10 (0·07)
GP_{18}	0·89 (0·02)	0·74 (0·33)	0·81 (0·02)	0·66 (0·02)	0·89 (0·04)	0·78 (0·03)	-0·05 (0·03)	1·09 (0·11)
GP_{15}	0·87 (0·02)	0·72 (0·07)	0·80 (0·02)	0·65 (0·04)	0·90 (0·06)	0·75 (0·04)	-0·05 (0·02)	1·09 (0·11)

6.A.3 *Feature importance of the GPx models*

For each GPx model, feature importance was calculated with the permutation importance inspection technique [38]. This methodology provides as a metric of feature importance the percent increase in misclassification rate when the value of a k^{th} feature are randomly shuffled. Shuffling the feature breaks the relationship between the feature and the target, therefore the consequent drop in performance is indicative of how much the model depends on that specific feature. Tables 6.A.4 to 6.A.11 report the results of the features importance analysis for the models GP_{34}, GP_{28}, GP_{26}, GP_{24}, GP_{22}, GP_{20}, GP_{18}, GP_{15} respectively. In the tables, the importance represents the percent increase in misclassification rate that follows the shuffling of the variable. The rank is calculated based on the importance.

The analysis shows that the features that were extracted from the ICP signal most contribute to the prediction task, in particular the median ICP value of the 10 minutes preceding the prediction, which is 40 minutes before the event to predict, had the highest predictive power in several models. Although features summarizing the median ICP value for different time windows presented high predictive power, an important role was equally played by more complex summary statistics such as standard deviations or frequency components analysis. Features extracted from the MAP and LAx signals had a minor role in the prediction, but were still selected from the features selection algorithm for almost all the GPx models.

Table 6.A.4 Feature importance of the model GP_{34}

Feature	Rank	Importance
ICP median minutes [230 to 240]	1st	0.01085
ICP SD minutes [60 to 120]	2nd	0.00513
ICP SD minutes [130 to 140]	3rd	0.00418
ICP 2nd largest Fourier coefficient	4th	0.00417
ICP 5th largest Fourier coefficient	5th	0.00259
ICP 1st cepstrum coefficient	6th	0.00196
ICP 3rd largest Fourier coefficient	7th	0.00180
ICP median minutes [210 to 240]	8th	0.00169
ICP median minutes [170 to 180]	9th	0.00139
ICP median minutes [90 to 100]	10th	0.00129
ICP median minutes [180 to 190]	11th	0.00126
ICP median minutes [160 to 170]	12th	0.00113
ICP median minutes [30 to 60]	13th	0.00089
ICP median minutes [140 to 150]	14th	0.00089
ICP SD minutes [0 to 10]	15th	0.00086
ICP median minutes [80 to 190]	16th	0.00081
ICP 1st largest Fourier coefficient	17th	0.00079
ICP 4th largest Fourier coefficient	18th	0.00076
ICP median minutes [220 to 230]	19th	0.00065
ICP median minutes [190 to 200]	20th	0.00059
ICP median minutes [150 to 180]	21st	0.00055
ICP median minutes [210 to 220]	22nd	0.00045

Table 6.A.5 Feature importance of the model GP_{28}

Feature	Rank	Importance
ICP median minutes [230 to 240]	1st	0.00725
ICP median minutes [210 to 240]	2nd	0.00318
ICP SD minutes [0 to 10]	3rd	0.00187
ICP median minutes [180 to 240]	4th	0.00185
ICP 5th largest Fourier coefficient	5th	0.00171
ICP median minutes [220 to 230]	6th	0.00167
LAx SD minutes [210 to 220]	7th	0.00158
ICP SD minutes [40 to 50]	8th	0.00144
ICP SD minutes [150 to 160]	9th	0.00143
MAP SD minutes [30 to 40]	10th	0.00140
LAx 2nd cepstrum coefficient	11th	0.00133
ICP SD minutes [90 to 120]	12th	0.00132
ICP median minutes [190 to 200]	13th	0.00131
ICP SD minutes [120 to 150]	14th	0.00125
Number of events of ICP dose in the red-subarea	15th	0.00121
ICP 2nd cepstrum coefficient	16th	0.00119
LAx median minutes [120 to 130]	17th	0.00115
MAP SD minutes [120 to 180]	18th	0.00113
ICP SD minutes [20 to 30]	19th	0.00111
LAx median minutes [210 to 220]	20th	0.00110
ICP SD minutes [60 to 70]	21st	0.00107
ICP SD minutes [10 to 20]	22nd	0.00102
ICP SD minutes [80 to 90]	23rd	0.00101
ICP median minutes [110 to 120]	24th	0.00098
ICP 1st largest Fourier coefficient	25th	0.00096
ICP median minutes [150 to 160]	26th	0.00094
LAx 3rd cepstrum coefficient	27th	0.00094
ICP SD minutes [60 to 120]	28th	0.00092
ICP SD minutes [170 to 180]	29th	0.00091
ICP median minutes [120 to 180]	30th	0.00090
ICP SD minutes [100 to 110]	31st	0.00088
LAx frequency 3rd largest Fourier coefficient	32nd	0.00088
ICP 1st cepstrum coefficient	33rd	0.00084
MAP SD minutes [220 to 230]	34th	0.00084
ICP median minutes [120 to 150]	35th	0.00081
ICP SD minutes [60 to 90]	36th	0.00081
ICP SD minutes [150 to 180]	37th	0.00079
LAx median minutes [170 to 180]	38th	0.00078
ICP median minutes [170 to 180]	39th	0.00074
ICP SD minutes [50 to 60]	40th	0.00071
ICP median minutes [90 to 120]	41st	0.00066
ICP SD minutes [140 to 150]	42nd	0.00064
ICP median minutes [100 to 110]	43rd	0.00062
LAx median minutes [130 to 140]	44th	0.00060
ICP SD minutes [110 to 120]	45th	0.00054
ICP median minutes [160 to 170]	46th	0.00050
Presence events of ICP dose in the red-subarea	47th	0.00050
ICP median minutes [60 to 170]	48th	0.00049
ICP median minutes [150 to 180]	49th	0.00048

Table 6.A.6 Feature importance of the model GP_{26}

Feature	Rank	Importance
ICP median minutes [230 to 240]	1st	0.01021
ICP median minutes [210 to 240]	2nd	0.00345
ICP median minutes [210 to 220]	3rd	0.00327
ICP SD minutes [180 to 210]	4th	0.00262
ICP median minutes [90 to 100]	5th	0.00205
Presence of “almost events” in the red-subarea	6th	0.00193
ICP SD minutes [60 to 70]	7th	0.00154
LAx median minutes [150 to 180]	8th	0.00147
Presence events of ICP dose in the red-subarea	9th	0.00146
ICP SD minutes [150 to 180]	10th	0.00145
ICP 5th largest Fourier coefficient	11th	0.00131
ICP SD minutes [140 to 150]	12th	0.00130
ICP median minutes [220 to 230]	13th	0.00126
ICP SD minutes [120 to 150]	14th	0.00108
ICP SD minutes [170 to 180]	15th	0.00099
LAx median minutes [160 to 170]	16th	0.00099
MAP median minutes [220 to 230]	17th	0.00098
ICP median minutes [180 to 240]	18th	0.00097
ICP SD minutes [0 to 30]	19th	0.00090
ICP 1st largest Fourier coefficient	20th	0.00079
ICP median minutes [160 to 170]	21st	0.00079
MAP median minutes [120 to 130]	22nd	0.00075
ICP median minutes [170 to 180]	23rd	0.00061
ICP SD minutes [130 to 140]	24th	0.00061
ICP median minutes [150 to 180]	25th	0.00058
ICP median minutes [120 to 180]	26th	0.00057
ICP median minutes [120 to 150]	27th	0.00056
ICP median minutes [140 to 150]	28th	0.00048
ICP median minutes [200 to 210]	29th	0.00043

Table 6.A.7 Feature importance of the model GP_{24}

Feature	Rank	Importance
ICP median minutes [230 to 240]	1st	0.00608
ICP median minutes [210 to 240]	2nd	0.00272
ICP median minutes [210 to 220]	3rd	0.00179
ICP median minutes [220 to 230]	4th	0.00167
ICP median minutes [180 to 240]	5th	0.00166
ICP 5th largest Fourier coefficient	6th	0.00158
ICP SD minutes [40 to 50]	7th	0.00140
Number of events of ICP dose in the red-subarea	8th	0.00135
ICP median minutes [170 to 180]	9th	0.00133
ICP SD minutes [120 to 180]	10th	0.00114
ICP median minutes [180 to 210]	11th	0.00110
ICP SD minutes [150 to 180]	12th	0.00103
LAx median minutes [180 to 240]	13th	0.00101
MAP median minutes [220 to 230]	14th	0.00097
MAP SD minutes [230 to 240]	15th	0.00096
ICP median minutes [120 to 180]	16th	0.00094
MAP SD minutes [40 to 50]	17th	0.00092
ICP SD minutes [60 to 90]	18th	0.00088
ICP 4th cepstrum coefficient	19th	0.00086
ICP median minutes [120 to 150]	20th	0.00085
ICP SD minutes [100 to 110]	21st	0.00085
ICP median minutes [110 to 120]	22nd	0.00084
ICP SD minutes [130 to 140]	23rd	0.00084
LAx SD minutes [170 to 180]	24th	0.00083
LAx SD minutes [120 to 150]	25th	0.00083
Presence events of ICP dose in the red-subarea	26th	0.00082
ICP SD minutes [180 to 190]	27th	0.00082
LAx median minutes [200 to 210]	28th	0.00082
ICP 1st largest Fourier coefficient	29th	0.00080
ICP SD minutes [50 to 60]	30th	0.00075
ICP SD minutes [120 to 130]	31st	0.00072
ICP median minutes [130 to 140]	32nd	0.00070
ICP SD minutes [210 to 220]	33rd	0.00069
ICP SD minutes [60 to 70]	34th	0.00068
ICP median minutes [150 to 180]	35th	0.00067
ICP 3rd cepstrum coefficient	36th	0.00065
LAx 5th cepstrum coefficient	37th	0.00058
ICP median minutes [140 to 150]	38th	0.00053
ICP median minutes [60 to 90]	39th	0.00051
ICP SD minutes [30 to 60]	40th	0.00050
ICP SD minutes [170 to 180]	41st	0.00042
ICP SD minutes [20 to 30]	42nd	0.00035

Table 6.A.8 Feature importance of the model GP_{22}

Feature	Rank	Importance
ICP median minutes [230 to 240]	1st	0.0060
ICP SD minutes [120 to 150]	2nd	0.0033
ICP median minutes [220 to 230]	3rd	0.0027
ICP median minutes [210 to 240]	4th	0.0024
ICP median minutes [130 to 140]	5th	0.0024
ICP median minutes [70 to 180]	6th	0.0023
ICP 4th largest Fourier coefficient	7th	0.0020
ICP median minutes [180 to 240]	8th	0.0019
ICP 4th cepstrum coefficient	9th	0.0019
ICP median minutes [110 to 120]	10th	0.0018
ICP SD minutes [180 to 190]	11th	0.0017
ICP SD minutes [40 to 50]	12th	0.0016
ICP median minutes [170 to 180]	13th	0.0016
ICP 5th largest Fourier coefficient	14th	0.0016
MAP SD minutes [120 to 180]	15th	0.0015
LAx median minutes [180 to 240]	16th	0.0014
Presence events of ICP dose in the red-subarea	17th	0.0012
ICP median minutes [180 to 190]	18th	0.0012
ICP median minutes [120 to 180]	19th	0.0012
ICP median minutes [120 to 150]	20th	0.0012
ICP median minutes [150 to 180]	21st	0.0010
MAP 2nd largest Fourier coefficient	22nd	0.0010
ICP median minutes [60 to 90]	23rd	0.0010
ICP SD minutes [200 to 210]	24th	0.0010
ICP 1st largest Fourier coefficient	25th	0.0010
ICP SD minutes [30 to 60]	26th	0.0008
ICP SD minutes [60 to 70]	27th	0.0008
ICP median minutes [180 to 210]	28th	0.0008
ICP median minutes [100 to 110]	29th	0.0008
ICP SD minutes [140 to 150]	30th	0.0007
ICP median minutes [60 to 170]	31st	0.0006

Table 6.A.9 Feature importance of the model GP_{20}

Feature	Rank	Importance
ICP median minutes [230 to 240]	1st	0.01053
ICP median minutes [210 to 240]	2nd	0.00357
MAP median minutes [220 to 230]	3rd	0.00319
ICP median minutes [170 to 180]	4th	0.00307
ICP median minutes [220 to 230]	5th	0.00271
ICP median minutes [180 to 240]	6th	0.00259
ICP 4th cepstrum coefficient	7th	0.00241
ICP median minutes [150 to 160]	8th	0.00240
ICP median minutes [120 to 150]	9th	0.00215
Number of events of ICP dose in the red-subarea	10th	0.00206
ICP median minutes [200 to 210]	11th	0.00203
ICP median minutes [180 to 190]	12th	0.00197
ICP median minutes [140 to 150]	13th	0.00191
ICP median minutes [160 to 170]	14th	0.00188
ICP median minutes [150 to 180]	15th	0.00180
ICP median minutes [120 to 180]	16th	0.00178
ICP SD minutes [40 to 50]	17th	0.00170
ICP SD minutes [20 to 30]	18th	0.00168
Presence events of ICP dose in the red-subarea	19th	0.00152
LAx SD minutes [180 to 210]	20th	0.00149
ICP SD minutes [60 to 90]	21st	0.00146
ICP median minutes [180 to 210]	22nd	0.00144
ICP 3rd cepstrum coefficient	23rd	0.00141
ICP SD minutes [70 to 80]	24th	0.00138
ICP median minutes [190 to 200]	25th	0.00135
MAP SD minutes [220 to 230]	26th	0.00132
ICP SD minutes [80 to 90]	27th	0.00112
Average duration of events of ICP dose in the red-subarea	28th	0.00109

Table 6.A.10 Feature importance of the model GP_{18}

Feature	Rank	Importance
ICP median minutes [50 to 160]	1st	0.00699
ICP median minutes [230 to 240]	2nd	0.00652
ICP median minutes [210 to 240]	3rd	0.00419
Average duration of events of ICP dose in the red-subarea	4th	0.00414
ICP median minutes [220 to 230]	5th	0.00383
ICP median minutes [120 to 130]	6th	0.00373
ICP median minutes [120 to 180]	7th	0.00341
ICP median minutes [180 to 240]	8th	0.00311
ICP median minutes [180 to 190]	9th	0.00302
ICP median minutes [130 to 140]	10th	0.00294
ICP median minutes [170 to 180]	11th	0.00276
ICP median minutes [60 to 120]	12th	0.00253
Number of “almost events” in the red-subarea	13th	0.00251
ICP 1st largest Fourier coefficient	14th	0.00244
ICP median minutes [120 to 150]	15th	0.00231
ICP median minutes [160 to 170]	16th	0.00219
LAx SD minutes [180 to 240]	17th	0.00183
ICP median minutes [150 to 160]	18th	0.00179
ICP median minutes [180 to 210]	19th	0.00160
ICP median minutes [190 to 200]	20th	0.00145
ICP median minutes [150 to 180]	21st	0.00133
ICP median minutes [90 to 120]	22nd	0.00130

Table 6.A.11 Feature importance of the model GP_{15}

Feature	Rank	Importance
Average duration of events of ICP dose in the red-subarea	1st	0.00964
ICP 3rd largest Fourier coefficient	2nd	0.00590
ICP median minutes [230 to 240]	3rd	0.00421
ICP 2nd cepstrum coefficient	4th	0.00363
ICP median minutes [210 to 220]	5th	0.00339
MAP 2nd cepstrum coefficient	6th	0.00290
ICP median minutes [210 to 240]	7th	0.00259
MAP 1st cepstrum coefficient	8th	0.00233
ICP median minutes [200 to 210]	9th	0.00218
Percentage of time with passive autoregulation	10th	0.00218
ICP median minutes [130 to 140]	11th	0.00214
Number of “almost events” in the red-subarea	12th	0.00206
ICP median minutes [20 to 130]	13th	0.00203
ICP median minutes [220 to 230]	14th	0.00200
ICP median minutes [180 to 240]	15th	0.00190
LAx 2nd largest Fourier coefficient	16th	0.00180
ICP 1st largest Fourier coefficient	17th	0.00179
LAx median minutes [160 to 170]	18th	0.00169
MAP SD minutes [230 to 240]	19th	0.00166
ICP median minutes [190 to 200]	20th	0.00165
ICP median minutes [60 to 120]	21st	0.00161
MAP 2nd largest Fourier coefficient	22nd	0.00159
ICP median minutes [40 to 150]	23rd	0.00153
LAx 4th cepstrum coefficient	24th	0.00151
ICP median minutes [180 to 210]	25th	0.00140
LAx 1st cepstrum coefficient	26th	0.00137
ICP 3rd cepstrum coefficient	27th	0.00136
LAx median minutes [200 to 210]	28th	0.00130
ICP median minutes [180 to 190]	29th	0.00127
LAx median minutes [190 to 200]	30th	0.00125
ICP SD minutes [50 to 60]	31st	0.00115
ICP median minutes [120 to 180]	32nd	0.00110
ICP median minutes [160 to 170]	33rd	0.00076

6.A.4 *Performance of the RFred model for the prediction of the red area*

The development cohort included 6959 events of ICP dose in the red area and 13979 randomly selected events in the blue area. Patients experienced on average a median [IQR] number of events of ICP dose in the red area of 7 [0 - 30]. In addition, they experienced a median [IQR] percentage of monitoring time spent in the red area of 8 [0-36], against a median [IQR] percentage of monitoring time spent with an ICP above 22 mmHg of 1 [0-6].

The model presented on the 10 folds CV internal validation sub-sets a mean (SD) AUC of 0·92 (0·02), a mean (SD) AP of 0·87 (0·03), a mean (SD) accuracy of 0·86 (0·02), a mean (SD) precision of 0·81 (0·04), a mean (SD) sensitivity of 0·76 (0·04), a mean (SD) specificity of 0·91 (0·02). The mean (SD) calibration-in-the-large and the mean (SD) calibration slope were respectively equal to 0·01 (0·06) and 0·92 (0·02). Visually, the RFred model presented good calibration (despite a p-value < 0·01) and superior clinical utility as compared to the “alert for all”/ “alert for none” options, see Figure 6.A.1.

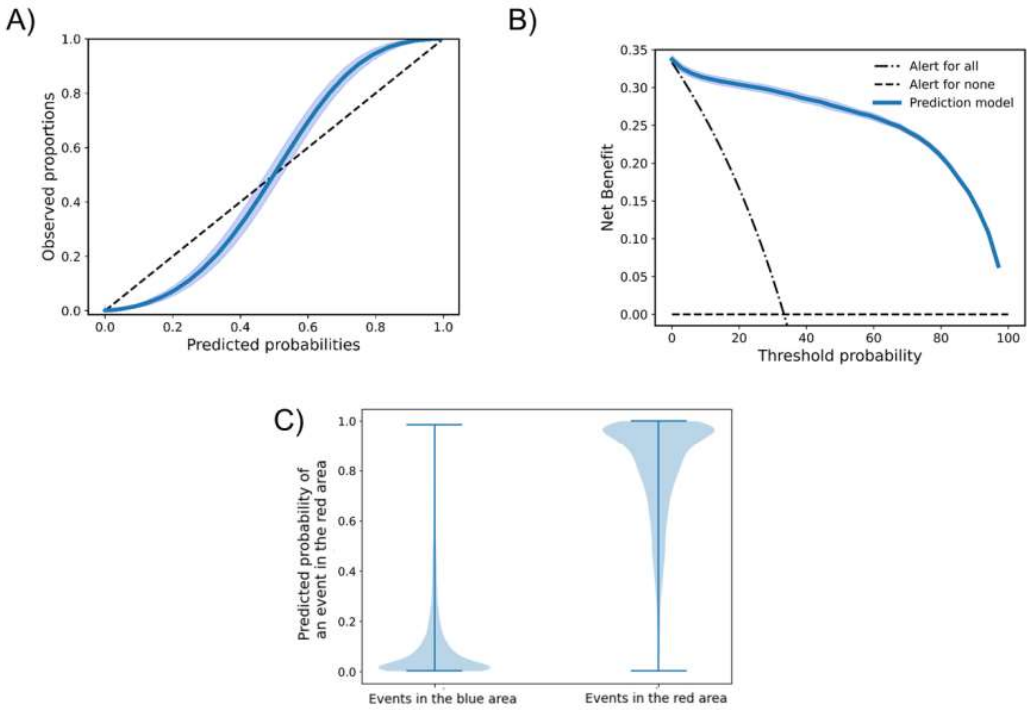


Figure 6.A.1 Performance metrics of the RFred model for the prediction of the red area on the development dataset. Panel A) calibration curve. Panel B) decision curve, for this model intervention may indicate the need of additional clinical evaluation. Panel C) violin plots of the predicted probabilities of an event in the red area or a not an event in relationship with the true labels. In the plot, the distributions of the probability scores of the false positive, true positive, true negatives and false negatives predictions are displayed.

6.A.5 *Calibration and decision curves of the GPx models*

Figure 6.A.2 below shows the calibration and decision curves of the GPx models when extracted on the development and on the CENTER-TBI dataset.

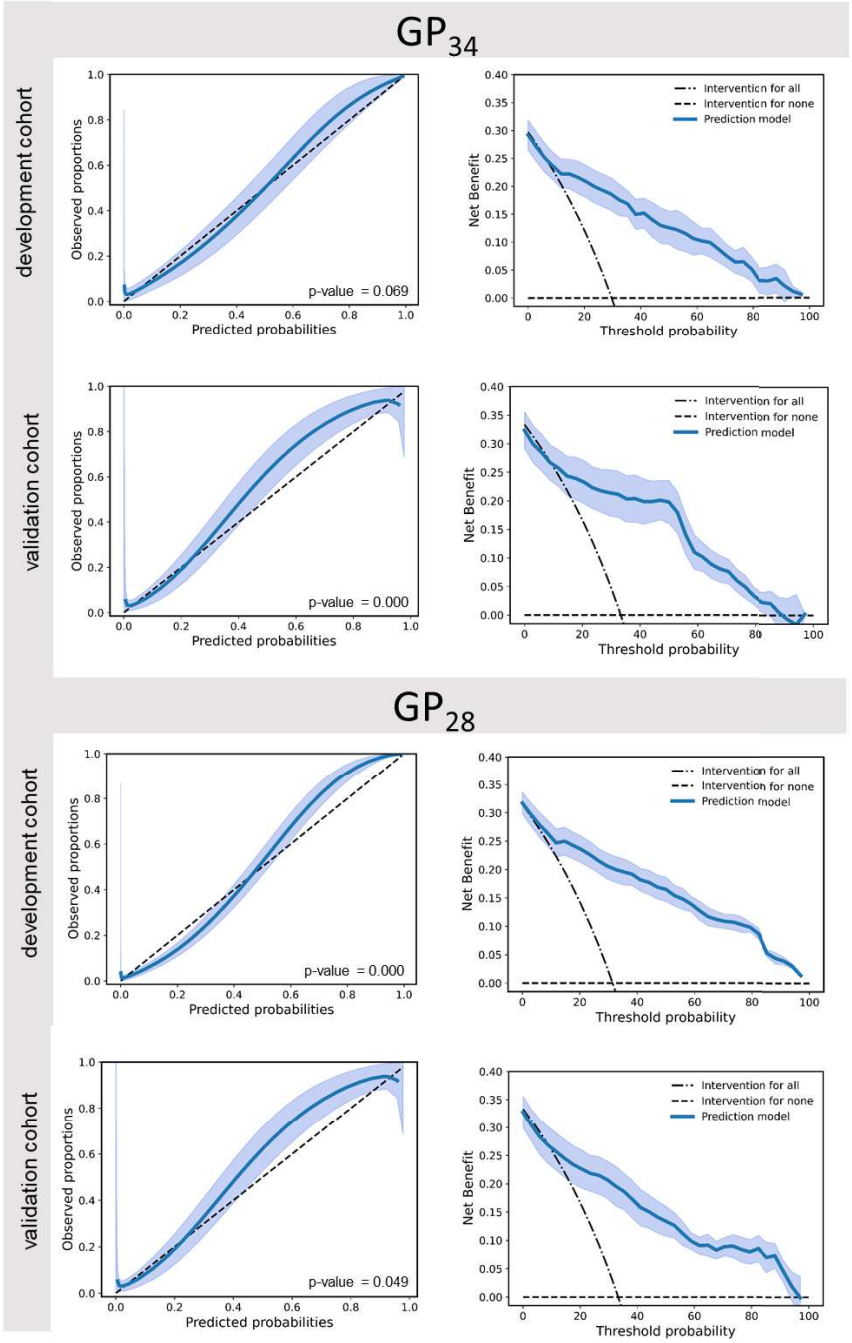


Figure 6.A.2 (part I) Calibrations and decision curves of the models GP_{34} and GP_{28} for the development and CENTER-TBI (external validation) cohorts. Calibration curves are displayed in the left column while decision curves on the right columns. The p-value for calibration is reported in each graph of calibration.

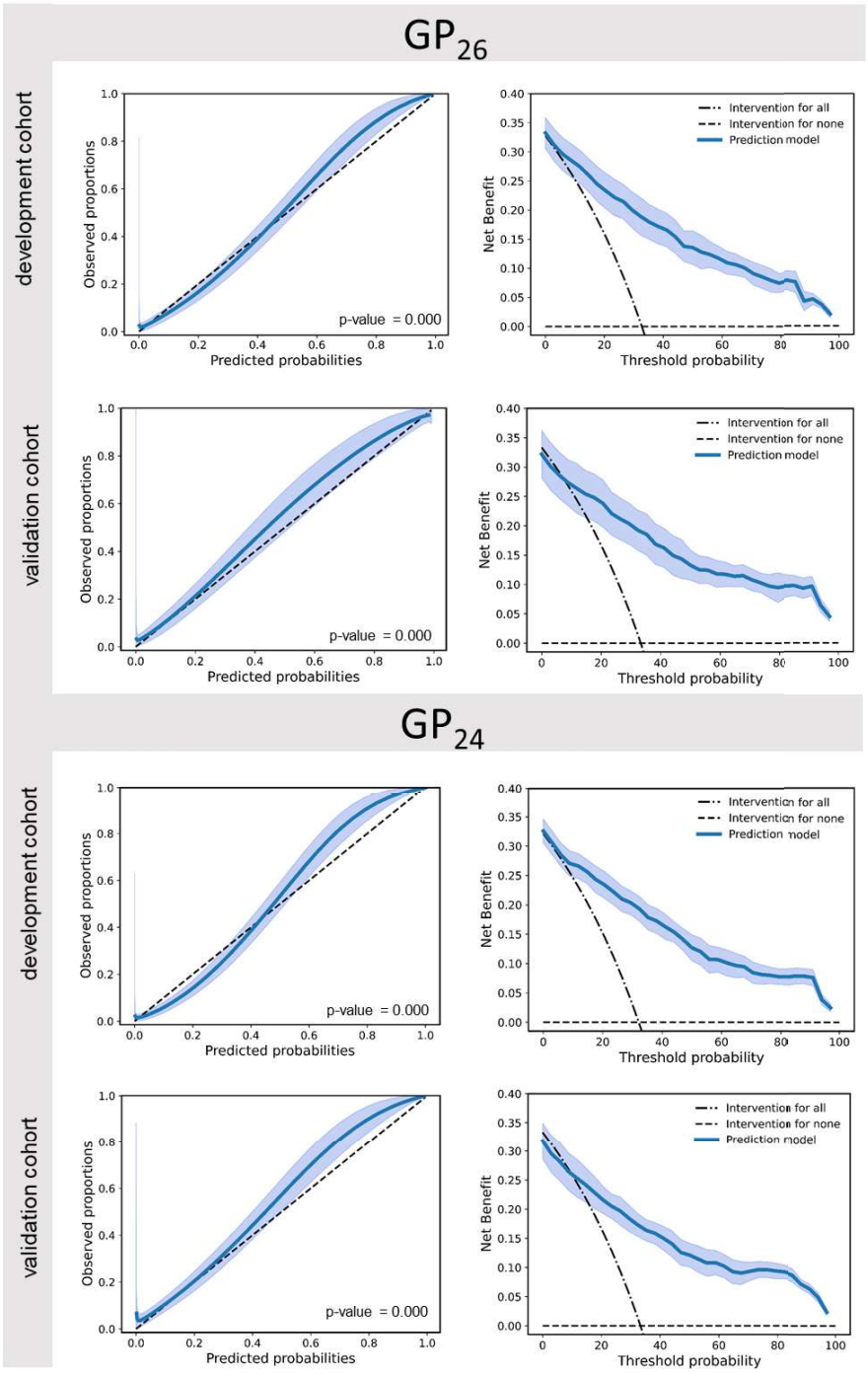


Figure 6.A.2 (part II) Calibrations and decision curves of the models GP_{26} and GP_{24} for the development and CENTER-TBI (external validation) cohorts. Calibration curves are displayed in the left column while decision curves on the right columns. The p-value for calibration is reported in each graph of calibration.

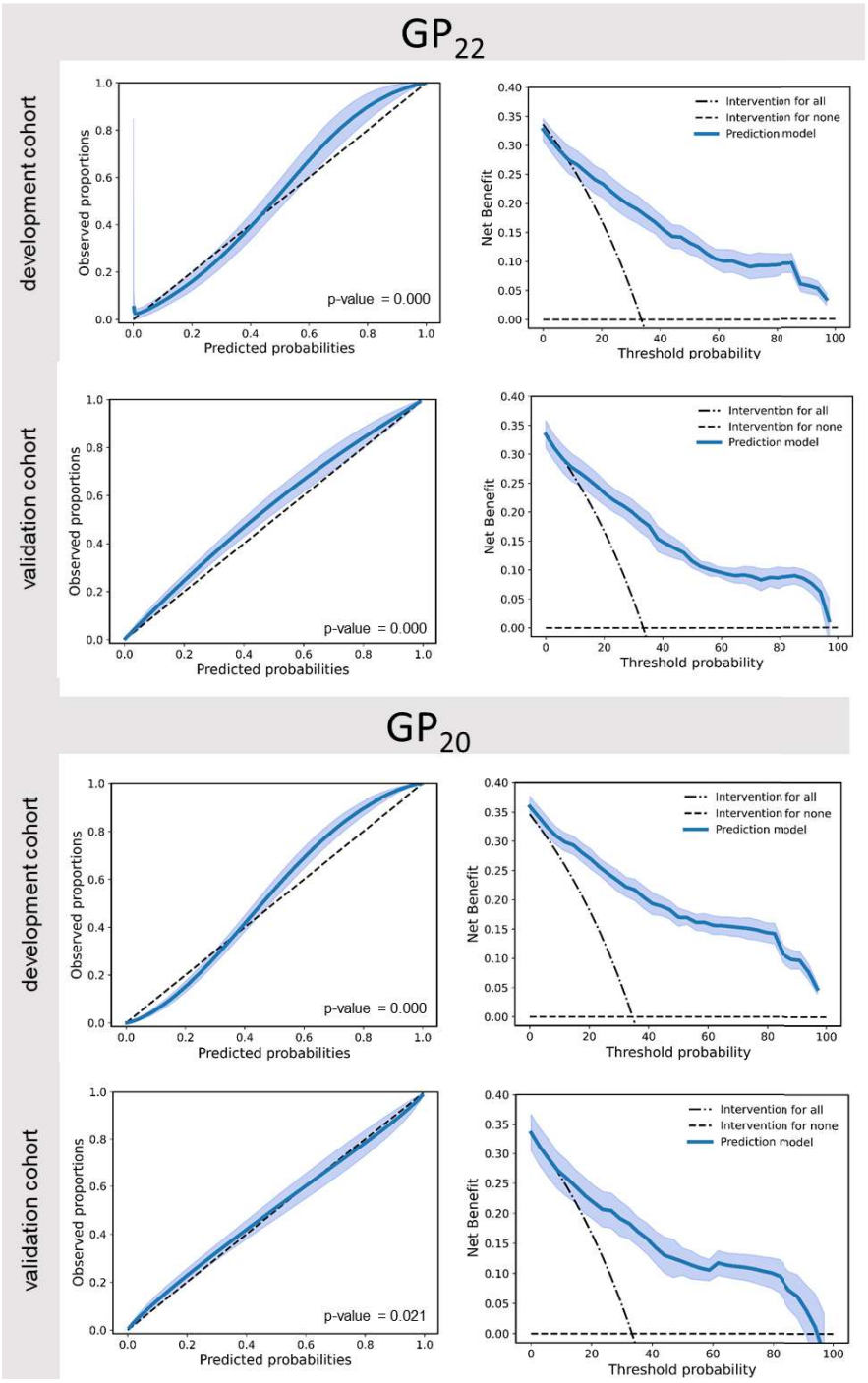


Figure 6.A.2 (part III) Calibrations and decision curves of the models GP_{22} and GP_{20} for the development and CENTER-TBI (external validation) cohorts. Calibration curves are displayed in the left column while decision curves on the right columns. The p-value for calibration is reported in each graph of calibration.

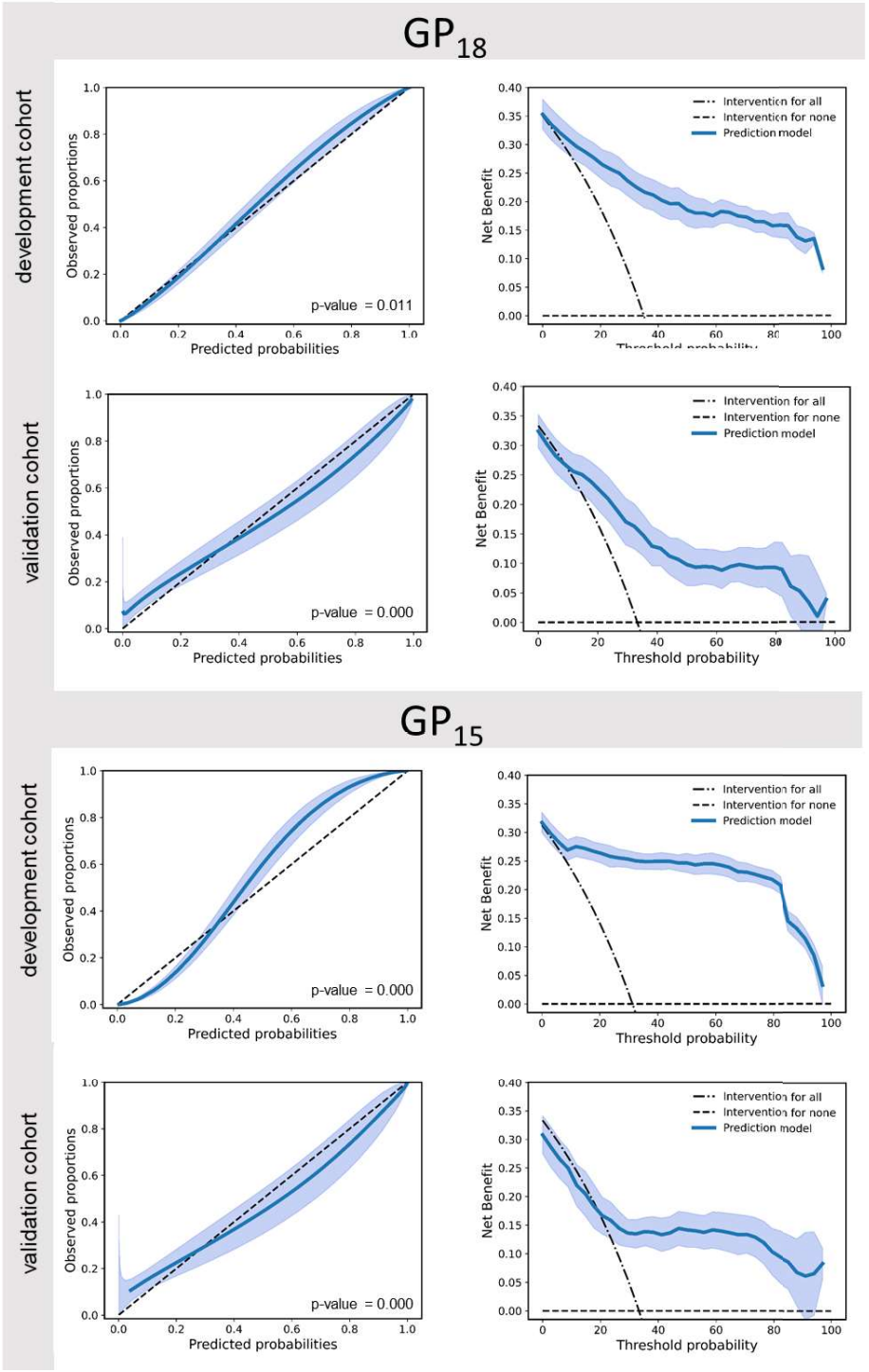


Figure 6.A.2 (part IV) Calibrations and decision curves of the models GP_{18} and GP_{15} for the development and CENTER-TBI (external validation) cohorts. Calibration curves are displayed in the left column while decision curves on the right columns. The p-value for calibration is reported in each graph of calibration.

6.A.6 *Performance of the model for different forewarning windows*

As an exploratory analysis we tested the performance of the model on the CENTER-TBI dataset (external validation dataset) when using different forewarning windows. In details, we tested the forewarning time of 5 minutes, 15 minutes, 45 minutes and 60 minutes. This exploratory analysis is hypothesis generating and could help to understand the performance of the model when used in a real time setting, where a strict forewarning time of 30 minutes may not apply.

Table 6.A.12 reports the performance of the model when tested for different forewarning windows. To facilitate the comparison, the Table 6.A.11 also reports the performance of the model for the prediction of the red area with a 30 minutes forewarning window (reference model presented in the main manuscript).

Table 6.A.12 Performance of the model when tested for different forewarning windows.

FOREWARNING WINDOW	AUC	AP	ACCURACY	PRECISION	SENSITIVITY	SPECIFICITY
5 minutes	0.94	0.91	0.89	0.86	0.79	0.94
15 minutes	0.94	0.88	0.89	0.82	0.78	0.93
30 minutes	0·94	0·89	0·88	0·85	0·77	0·93
45 minutes	0.93	0.89	0.88	0.85	0.77	0.93
60 minutes	0.93	0.89	0.88	0.85	0.76	0.93

BIBLIOGRAPHY

1. Stocchetti N, Maas AIR (2014) Traumatic Intracranial Hypertension. N Engl J Med 370:2121–2130. https://doi.org/10.1056/NEJMra1208708

2. Maas AIR, Menon DK, Adelson PD, et al (2017) Traumatic brain injury: integrated approaches to improve prevention, clinical care, and research. Lancet Neurol 4422:. https://doi.org/10.1016/S1474-4422(17)30371-X

3. Marmarou A, Anderson RL, Ward JD, et al (1991) Impact of ICP instability and hypotension on outcome in patients with severe head trauma. J Neurosurg 75:S59–S66. https://doi.org/10.3171/sup.1991.75.1s.0s59

4. Stocchetti N, Zanaboni C, Colombo A, et al (2008) Refractory intracranial hypertension and "second-tier" therapies in traumatic brain injury. Intensive Care Med 34:461–467. https://doi.org/10.1007/s00134-007-0948-9

5. Carney N, Totten AM, O'Reilly C, et al (2016) Guidelines for the Management of Severe Traumatic Brain Injury 4th Edition

6. Meyfroidt G, Citerio G (2017) Letter: Guidelines for the Management of Severe Traumatic Brain Injury, Fourth Edition. Neurosurgery 81:E1–E1. https://doi.org/10.1093/neuros/nyx144

7. Helbok R, Meyfroidt G, Beer R (2018) Intracranial pressure thresholds in severe traumatic brain injury: Con: The injured brain is not aware of ICP thresholds! Intensive Care Med 44:1318–1320. https://doi.org/10.1007/s00134-018-5249-y

8. Stocchetti N, Poole D, Okonkwo DO (2018) Intracranial pressure thresholds in severe traumatic brain injury: we are not sure: Prudent clinical practice despite dogma or nihilism. Intensive Care Med 44:1321–1323. https://doi.org/10.1007/s00134-018-5251-4

9. Myburgh JA (2018) Intracranial pressure thresholds in severe traumatic brain injury: Pro. Intensive Care Med 44:1315–1317. https://doi.org/10.1007/s00134-018-5264-z

10. Sorrentino E, Diedler J, Kasprowicz M, et al (2012) Critical Thresholds for Cerebrovascular Reactivity After Traumatic Brain Injury. Neurocrit Care 16:258–266. https://doi.org/10.1007/s12028-011-9630-8

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

12. Donnelly J, Güiza F, Depreitere B, et al (2021) Visualising the pressure-time burden of elevated intracranial pressure after severe traumatic brain injury: a retrospective confirmatory study. Br J Anaesth 126:e15–e17. https://doi.org/10.1016/j.bja.2020.09.018

13. Vik A, Nag T, Fredriksli O, et al (2008) Relationship of " dose " of intracranial hypertension to outcome in severe traumatic brain injury. J Neurosurg J Neurosurg J Neurosurg 109:0–678. https://doi.org/10.3171/JNS/2008/109/10/0678

14. Lazaridis C, DeSantis SM, Smielewski P, et al (2014) Patient-specific thresholds of intracranial pressure in severe traumatic brain injury. J Neurosurg 120:893–900. https://doi.org/10.3171/2014.1.JNS131292

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

16. Chambers IR (2006) Critical thresholds of intracranial pressure and cerebral perfusion pressure related to age in paediatric head injury. J Neurol Neurosurg Psychiatry 77:234–240. https://doi.org/10.1136/jnnp.2005.072215

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

18. Jennett B, Bond M, B.Jennett; M.Bond (1975) Assessment of Outcome After Severe Brain Damage. Lancet 1:480–484

19. Myers RB, Lazaridis C, Jermaine CM, et al (2016) Predicting Intracranial Pressure and Brain Tissue Oxygen Crises in Patients with Severe Traumatic Brain Injury. Crit Care Med 44:1754–1761. https://doi.org/10.1097/CCM.0000000000001838

20. Scalzo F, Hamilton R, Asgari S, et al (2012) Intracranial hypertension prediction using extremely randomized decision trees. Med Eng Phys 34:1058–1065. https://doi.org/10.1016/j.medengphy.2011.11.010

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

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

23. Güiza F, Depreitere B, Piper I, et al (2013) Novel Methods to Predict Increased Intracranial Pressure During Intensive Care and Long-Term Neurologic Outcome After Traumatic Brain Injury: Development and Validation in a Multicenter Dataset. Crit Care Med 41:554–564. https://doi.org/DOI 10.1097/CCM.0b013e3182742d0a

24. Olson DM, Andrew Kofke W, O'Phelan K, et al (2015) Global Monitoring in the Neurocritical Care Unit. Neurocrit Care 22:337–347.

https://doi.org/10.1007/s12028-015-0132-y

25. Le Roux P (2016) Intracranial Pressure Monitoring and Management. In: Laskowitz D, Grant G (eds) Translational Research in Traumatic Brain Injury. CRC Press/Taylor and Francis Group, Boca Raton (FL)

26. Depreitere B, Citerio G, Smith M, et al (2021) Cerebrovascular Autoregulation Monitoring in the Management of Adult Severe Traumatic Brain Injury: A Delphi Consensus of Clinicians. Neurocrit Care 34:731–738. https://doi.org/10.1007/s12028-020-01185-x

27. Czosnyka M, Smielewski P, Kirkpatrick P, et al (1997) Continuous assessment of the cerebral vasomotor reactivity in head injury. Neurosurgery 41:11–7; discussion 17-9

28. Depreitere B, Güiza F, Van den Berghe G, et al (2014) Pressure autoregulation monitoring and cerebral perfusion pressure target recommendation in patients with severe traumatic brain injury based on minute-by-minute monitoring data. J Neurosurg 120:1451–1457. https://doi.org/10.3171/2014.3.jns131500

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

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

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

32. (2007) Guidelines for the management of severe traumatic brain injury (3rd edition). J Neurotrauma 24:

33. Carra G, Güiza F, Depreitere B, Meyfroidt G (2020) Prediction model for intracranial hypertension demonstrates robust performance during external validation on the CENTER-TBI dataset. Intensive Care Med. https://doi.org/10.1007/s00134-020-06247-4

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

35. Tibshirani R (1996) Regression Shrinkage and Selection via the Lasso. J R Stat Soc 267–288

36. Frénay B, Doquire G, Verleysen M (2013) Is mutual information adequate

for feature selection in regression? Neural Networks 48:1–7. https://doi.org/10.1016/j.neunet.2013.07.003

37. Shannon CE (1948) A Mathematical Theory of Communication. Bell Syst Tech J 27:379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

38. Leo Breiman (2001) Random Forests. Mach Learn 45:5–32

39. Moons KGM, Altman DG, Reitsma JB, et al (2015) Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): Explanation and elaboration. Ann Intern Med 162:W1–W73. https://doi.org/10.7326/M14-0698

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

41. Güiza F, Van Eyck J, Meyfroidt G (2013) Predictive data mining on monitoring data from the intensive care unit. J Clin Monit Comput 27:449–453. https://doi.org/10.1007/s10877-012-9416-3

42. Leisman DE, Harhay MO, Lederer DJ, et al (2020) Development and Reporting of Prediction Models. Crit Care Med 48:623–633. https://doi.org/10.1097/CCM.0000000000004246

43. Flechet M, Meyfroidt G, Piper I, et al (2018) Visualizing Cerebrovascular Autoregulation Insults and Their Association with Outcome in Adult and Paediatric Traumatic Brain Injury. In: Acta Neurochirurgica Supplementum. pp 291–295

44. Depreitere B, Güiza F, Van den Berghe G, et al (2014) Pressure autoregulation monitoring and cerebral perfusion pressure target recommendation in patients with severe traumatic brain injury based on minute-by-minute monitoring data. J Neurosurg 120:1451–1457. https://doi.org/10.3171/2014.3.JNS131500

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

7

DEVELOPMENT AND TECHNICAL VALIDATION OF THE TB-AI: A PROTOTYPE SOFTWARE FOR THE MANAGEMENT OF ELEVATED INTRACRANIAL PRESSURE IN SEVERE TRAUMATIC BRAIN INJURY

ADAPTED FROM: **Carra G.** et al. “Development and technical validation of the TB-AI: a prototype software for the management of elevated intracranial pressure in severe traumatic brain injury”. Manuscript ready for submission.

ABSTRACT

OBJECTIVE: To develop a new decision-support software prototype for the management of intracranial hypertension in patients with severe traumatic brain injury (TBI) and to test such prototype at the bedside in a blinded mode. The software displays prediction of future events of elevated intracranial pressure (ICP) that could be potentially dangerous for the patient as well as results of previous research that could provide useful information to the clinicians.

APPROACH: We developed a prototype software for the management of patients with TBI in close collaboration with nurses and clinicians of the intensive care unit (ICU) of the University Hospitals of Leuven (UZ Leuven). We tested the functioning of the prototype in a real-time clinical setting in an observational technical validation. The study was performed in the ICU of Leuven between January 2020 and April 2021. Software defects that arose during the real-time use of the model were detected and solved, where possible.

MAIN RESULTS: The user interface of the software contains information that is potentially useful for the management of patients with TBI. During the technical validation we tested the software on 35 patients with continuous ICP monitoring. More that 450 tests on the correct functioning of the software were executed. Six software defects were identified and 5 of them were correctly solved.

SIGNIFICANCE: We developed a prototype software that could potentially support clinicians during the management of patients with TBI. The novelty of the software lies in the displayed predictions and metrics, which are potentially relevant in the clinical evaluation of patients with TBI, although not yet available at the bedside. The developed software could be used in future interventional studies that aim to assess the impact of recent results of TBI research on patients outcomes.

7.1 INTRODUCTION

In a neuro-intensive care unit (NICU), monitoring and treatment of intracranial pressure (ICP) is a key aspect in the management of patients with severe traumatic brain injury (TBI) [1]. Increased ICP levels are frequent after TBI and can lead to secondary brain damage (SBI) or death [1, 2]. To prevent SBI due to elevated ICP, current recommendations suggest to start aggressive treatment when the ICP rises above 22 mmHg [3]. However, a threshold-based approach may not deliver optimal management. The use of a universal threshold that is applied to every patient regardless of their clinical status and personal needs, may lead to missed treatment opportunities, or it may even trigger a too early administration of aggressive treatments, exposing the patient to unjustified risk.

The "dose" of ICP, which is a composite measure that takes into consideration the intensity and the duration of an ICP event [4], may better quantify the neurological burden related to elevated ICP. The association between the time-intensity ICP and outcomes was visualized by Güiza et al. [5] see Figure 7.1 panel A, and replicated in other two studies [6, 7]. The plot shows two distinct regions of ICP doses that are associated with good and poor long-term neurological outcomes. Specifically, the region in blue indicates ICP doses that occur more frequently in patients with better long-term outcomes, while the region in red indicates ICP doses that occur more frequently in patients with worse long term outcomes. Although the concept of ICP dose provides new insights on the burden of ICP in severe TBI and could potentially promote a more appropriate treatment for elevated ICP, until now no tool was available in the clinic to display this information.

Early warnings of potentially harmful episodes of ICP could result in a more appropriate and timely treatment of elevated ICP. For this purpose, Güiza et al. [8] developed a machine learning (ML) model for the prediction of extremely elevated ICP [9]. The model predicted with a 30 minutes forewarning events of ICP>30 mmHg that lasted more than 10 minutes. In a following study from Carra et al. (Chapter 6), this initial model was later integrated in a ML model that predicted with 30 minutes forewarning a larger set of episodes of ICP doses associated with poor outcomes, i.e. the red area of the visualization of Güiza et al.[5]. Both models base their predictions on continuous ICP and mean arterial blood pressure (MAP)

traces. The integration of such models at the bedside may have a positive effect on quality of care and patient outcomes.

Cerebrovascular autoregulation (CAR) represents the capacity of the brain to adjust the cerebral vessel resistance in response to changes in arterial blood pressure or cerebral perfusion, in order to ensure constant cerebral blood flow. In TBI, CAR is often disturbed, affecting the capacity of the brain to tolerate elevated ICP. A validated method to assess CAR is still missing, but CAR can be quantified with computation indexes such as the Pressure Reactivity index (PRx) [10] or the low-frequency autoregulatory index (LAx) [11], which can be extracted from the ICP and MAP signals. In a previous study, Flechet et al. [12] used the visualization method proposed by Güiza et al. [5] to display the association between doses of LAx and outcomes, see Figure 7.1 panel B. The study from Flechet et al. [12] demonstrated that also duration and intensity of LAx plays a role in the definition of the long-term neurocognitive outcomes of patients with TBI. No information about the CAR status of the patients is commercially available in the clinic. Only some specific research centers have access to continuous CAR status via a tool exclusively for research purposes (ICM+, Cambridge Enterprise Ltd., UK).

In order to assess the impact of the ICP dose and CAR dose visualizations on patient outcomes, it is important to develop bedside research tools that implement these results and display them in a relevant and interpretable fashion. These tools need to be robust and reliable and need to easily integrate with the clinical setting.

Specific goals of this study were 1) the design and implementation of a bedside software that integrates previous research on TBI in a unique bedside tool, and therefore provides to nurses and clinicians new information for the management of patients; 2) the design of a user friendly and interpretable interface in collaboration with nurses and clinicians; and 3) the technical validation of the software in a prospective observational study. The technical validation aims at detecting and correcting software defects that could rise during the use of the prototype software in real-time. Part of the data collected in this study were also used to assess the real-time performance of the prediction model for harmful ICP doses. The results of this performance analysis are reported separately in Chapter 8.

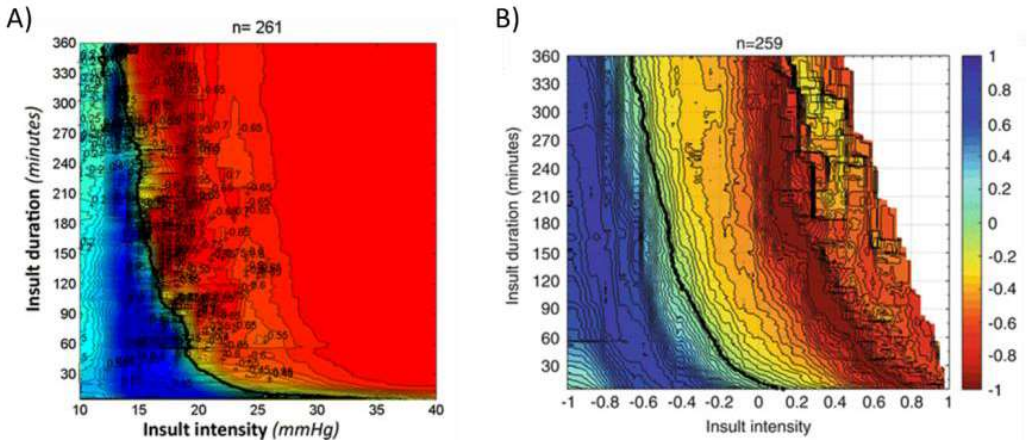


Figure 7.1 Visualization of the association between dose episodes and long-term neurological outcomes. Panel A) Visualization of the association between doses of ICP and long-term neurological outcomes. The red area indicates doses of ICP that occur more frequently in patients with poor long-term outcomes. The blue area indicates doses of ICP that occur more frequently in patients with better long-term outcomes. Panel B) Visualization of the association between doses of LAx and long-term neurological outcomes. Similarly, the red and blue areas indicate doses of LAx that occur more frequently in patients with poor and better long-term outcomes, respectively. Reproduced with permission from Güiza et al "Visualizing the pressure and time burden of intracranial hypertension in adult and paediatric traumatic brain injury" Springer Nature. And Flechet et al. "Visualizing Cerebrovascular Autoregulation Insults and Their Association with Outcome in Adult and Paediatric Traumatic Brain Injury" Springer Nature.

7.2 METHODS

7.2.1 *Study population*

The technical validation study was performed in the intensive care unit (ICU) of the University Hospitals UZ Leuven, Belgium, between January 2020 and April 2021 and included all patients that underwent intra-parenchymal ICP monitoring. Inclusion was not strictly limited to patients with TBI given the technical nature of the study. For the same reason, we did not perform a formal sample size calculation. Patients that were subjected to a therapy restriction code at the moment of screening were excluded from the study. During the technical validation the software was running in a blinded mode, therefore the standard clinical management of patients was not affected. The medical ethics committee of the University Hospitals UZ Leuven, Belgium approved this study. Patients and/or their

families were informed about the prospective collection of data though an information document.

7.2.2 *Software development*

The prototype software was called TB-AI, which stands for "Artificial Intelligence for Traumatic Brain Injury". The goal of the TB-AI was to implement 3 main sources of information from previous research on TBI that are not currently available at the bedside but could be potentially useful for the management of patients. Namely:

- Predictions of potentially harmful ICP elevations
- The LAx as a computational index of CAR
- Visualization of the real-time time-intensity burden of ICP and LAx

Each of these metrics is computed and displayed on a minute-by-minute basis. The TB-AI software was developed in Python (*version 3.5*, https://www.python.org/) with the Kivy library (*version 1.10*, https://kivy.org/#home) and it was based on a Model-View-Controller (MVC) architectural pattern, which separates the application into three main logical components: the model, the view and the controller. Each of these logical components handles a specific function of the application. For more information on the software architecture, see Appendix 7.A.1.

For practical reasons, the technical validation was performed by implementing the prediction model of Güiza et al. [8] for the prediction of ICP above 30 mmHg for more than 10 minutes. However, due to the MVC architecture this model is perfectly interchangeable with the model proposed by Carra et al. (Chapter 6) for the prediction of harmful ICP doses (given that input and output types are the same).

In the intensive care unit (ICU) of UZ Leuven, data of the patients are stored in the Patient Data Management System (PDMS) (MetaVision, iMD-Soft, Needham, MA,USA). The PDMS collects several clinical data of the patients, such as demographics, laboratory tests analysis, imaging data, notes from the clinical evaluations of the patient, treatment information and continuous monitoring data, which are stored on a minute-by-minute basis. Data are stored in a structured query language (SQL) database, which allows to handle big datasets of different data type with ease. The TB-AI software is connected to the UZ Leuven hospital network and it queries

through an encrypted connection the necessary data, i.e. continuous ICP and MAP signals, directly from the PDMS. This architecture allows the TB-AI software to process real-time minute-by-minute data of the patient. This is also a flexible architecture that is not monitor-dependent and that can be easily transferred to other hospitals that have an electronic health records management system in place. Moreover, to meet the safety regulations of the hospital an access log (record of all the actions done by each user) had to be implemented. In addition, access to the software could be done only by personal password.

7.2.3 *Technical validation*

The technical validation aimed at identifying software defects that could arise during the use of the software in a real-time clinical setting. Software defects were identified by performing the following tests at random moments every 2/3 monitoring days:

- Test on the correct acquisition and storage of the monitoring data
- Test on the computation of the LAx
- Test on the computation of the ICP predictions
- Test on the computation of the ICP statistics
- Test on the correct real-time visualization of the ICP and LAx burden plots
- Test on computational delays (for example potential delays in the computation of metrics, or visualization delays)

These tests aimed at identifying, among others, potential computational and visualization errors or errors in the acquisition or storage of data. For each identified software defects a defect report was completed, a solution was proposed, where possible, and the software was iteratively re-tested at the bedside.

7.2.4 *User interface*

The user interface (UI) of the software was designed in close collaboration with research nurses and clinicians of the ICU of UZ Leuven. In detail, we organized 3 meetings in which we presented the UI to a group of 6 nurses and clinicians and then we iteratively improved it based on their comments and suggestions.

7.3 RESULTS

7.3.1 Technical validation

Thirty-six patients with invasive intra-parenchymal ICP monitoring were included in this study, of which 14 patients had severe TBI, 11 patients had intracerebral hemorrhage, 6 patients had sub-arachnoid hemorrhage, 4 patients had meningitis and 1 patient had invasive ICP monitoring as a result of a brain tumor. Patients underwent invasive ICP monitoring for median [IQR] 7 [5-10] days.

During the technical validation, we performed more than 450 tests on the technical functioning of the model. Of these tests, 6 identified a software defect. A summary of the software defects is reported in Appendix 7.A.2. Importantly, no defects related to time delays or prediction calculation and visualization were identified. Moreover, the only defects that could not be solved concerned connection interruptions between the software and the PMDS. These interruptions occur during the weekly safety updates of the PDMS. Given that it was an error related to the hospital safety it could not be solved. Five times the software was interrupted due to human error, the interface was modified in order to minimize such problems.

7.3.2 User interface

The final version of the TBI-AI software UI presents three main screens: the home screen, the monitoring signals screen and the ICP and LAx burden visualization screen. Each screen is the result of a close collaboration with nurses and clinicians. Figure 7.2 shows a comparison between the initial UI of the TB-AI software, prior to any feedback session, and the final UI.

The home screen is divided into three panels. In the top panel, the predictions of harmful ICP events are displayed in a color-coded fashion. Red colors indicate higher probabilities of future ICP elevations while blue colors indicates lower probabilities of elevated ICP. In the middle panel, continuous traces of ICP and LAx are displayed. The bottom-left panel displays the percentage of time that the patients spent in the red area. It is possible to select the window of time in which this metric is computed among the following options: last 4 hours of monitoring, last 24 hours of monitoring or entire length of monitoring (LOM). The bottom-right panel displays the percentage of monitoring time that the patient spent with an ICP above a certain threshold, which can be chosen by the user among the

options of 20 mmHg, 25 mmHg or 30 mmHg. This metric can be similarly computed for different windows of time, i.e. last 4 hours of monitoring time, last 24 hours of monitoring time or entire LOM. All metrics are computed in real time and updated each minute. A zoomed version of Figure 1 panel B can be found in Appendix 7.A.3 (Figure 7.A.2).

The monitoring signals screen shows continuous minute-by-minute traces of relevant physiological signals for the management of patients with TBI. In detail, the screen shows the following signals: ICP, LAx, MAP, end-tidal CO_2 and the brain tissue oxygen ($PbtO_2$). This screen was not part of the initial design of the UI, but its importance was stressed during the feedback sessions. A zoomed version of Figure 1 panel C can be found in Appendix 7.A.3 (Figure 7.A.3).

The ICP and LAx burden visualization screen shows the real-time visualization plots of the experienced dose of ICP and LAx in relations with long-term outcomes. In detail, the left panel displays the real-time visualization of the ICP burden. The current ICP dose is plotted in vivid colors, while the cumulative ICP burden of the previous monitoring time, i.e. the worst ICP burden dose experienced by the patient, is plotted with more transparent colors. In the bottom left corner of the screen the percentage of time spent by the patient with an ICP in the “red area” of the visualization is displayed. The user can choose to have this percentage computed for different time windows, respectively for the past 4 hours, 24 hours or the LOM. In the right panel the real-time visualization of the LAx burden is displayed. The visualization follows the same principles of the ICP burden visualization plots explained before. Similarly to the left panel, the bottom right panel shows the percentage of time spent by the patient with a LAx in the “red area” of the visualization. As mentioned before, this percentage can be computed for different time windows. A zoomed version of Figure 1 panel D can be found in Appendix 7.A.3 (Figure 7.A.4).

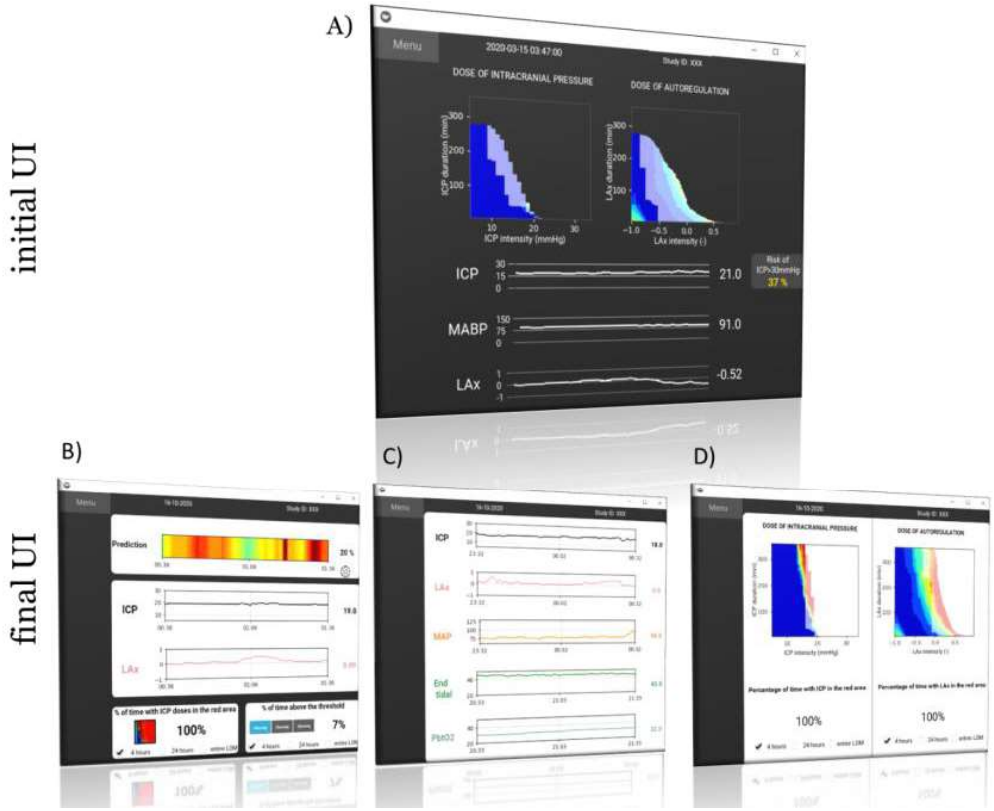


Figure 7.2 User Interface (UI) of the TB-AI software. Panel A initial UI of the software. This UI was used as starting point to integrate the suggestions and feedback of a group of 6 nurses and clinicians of the intensive care unit of UZ Leuven. Panel B, C and D show the final UI, which integrates the feedback obtained by the clinical personnel. In detail, panel B) home screen of the TBI-AI software, it shows the predictions of impending events of harmful intracranial pressure (ICP) so as relevant summary statistics on the current and previous ICP burden. Panel C) monitoring signals screen, it displays minute-by-minute traces of relevant physiological signals. Panel D) ICP and low-frequency autoregulatory index (LAx) burden visualization screen, it displays the visualization plots of the ICP and LAx burden, it also displays relevant metrics on the monitoring time that was spent by the patient with ICP or LAx in the red area of the visualization plots of Güiza et al. [5] and Flechet et al. [12].

7.4 DISCUSSION

Current management of severe TBI relies on a threshold-based approach, with several limitations. A redefinition of the type of elevated ICP events that might incur in harm and the reevaluation of ICP management towards a more personalized approach may improve patients' outcomes. Past

research has taken multiple steps in this direction, but none of the past research results is currently available at the bedside in a clinical tool.

In this study, we developed and validated a prototype software for the management of patients with TBI. The TB-AI prototype software displays potentially useful information not yet available at the bedside. This prototype may represent a valuable research tool to assess whether the use at the bedside of the concept of dose of ICP, continuous predictions of harmful ICP events or the LAx as an index of CAR may have an effect on management and outcomes in patients with TBI.

When developing tools for the use at the bedside, interdisciplinarity is key. Importantly, the UI of the software was designed in close collaboration between nurses and clinicians, with the objective of designing an interpretable tool that displays information in a clinically relevant fashion. Further assessment of the information displayed in the software needs to be performed once the software is used in clinical practice. During bedside use, unforeseen situations may arise for which the computation of additional metrics may be of interest. For example, in patients that undergo surgery for mass evacuation it may be of interest to re-start the computation of some metrics in relation to the moment of surgery. Moreover, additional tests on software usability need to be performed when the software is used in a non-blinded mode.

The developed prototype was tested in a clinical setting demonstrating its ability to compute the displayed metrics in real time. In addition, during the clinical test we investigated the correct functioning of the software and identified and solved 6 software defects. This technical validation guarantees accuracy and establishes the software as a reliable tool that could be used for a future prospective interventional study. Moreover, the technical validation provided us with valuable insights on the technical challenges that can be encountered when bringing a ML model to the bedside. Most of these technical challenges were related to the implementation and/or interaction with the network and software safety measures required by the local hospital. Open collaboration with the information technology (IT) department of the hospital is extremely important to solve or avoid these problems, therefore it is suggested to involve the IT department of the local hospitals from the very first phases of software design.

This prototype software provides additional information for the management of TBI, but potential changes in patient therapy remain at discretion of the clinician. In addition, it is important that clinicians are aware of the methodology behind the computation of the displayed information for a correct interpretation of the displayed results.

7.5 CONCLUSIONS

In this study we have developed and validated in a real clinical setting a prototype software for the management of patients with severe TBI. The software displays relevant metrics and information not currently available in the clinic and as such it could provide valuable additional information for the treatment of patients with severe TBI. The software could be a useful tool for setting up a future interventional study that assesses the positive impact of recent research results on patients management and outcomes.

ACKNOWLEDGMENT AND PERSONAL CONTRIBUTION

Study concept and design	Güiza, Meyfroidt, Flechet, **Carra,** Depreitere
Active role in the execution of the technical validation	**Carra,** Mebis, Wouters, Chao-Yuan Huang, Vanhulle, Hendrickx, Witpas, Dietvorst
Software implementation	**Carra,** Flechet
Design of the interface	**Carra,** Güiza, Meyfroidt, Depreitere, Flechet, Van den Berghe, Mebis, Wouters, Dietvorst
Drafting of the manuscript	**Carra**
Manuscript revision	**All authors**
Principal investigator	Meyfroidt

7.A APPENDIX

7.A.1 *Model-View-Controller architecture*

The Model-View-Controller (MVC) architectural pattern separates the application into three main logical components: the model, the view and the controller. Each of these logical components handles a specific function of the application: the model is the computational and data-related logic of the application, it executes the main operations of the application and it is involved in the storing and querying of the data. The view is the user interface (UI) logic of the application, it executes the operations able to visualize in an interpretable way the results and functionalities of the application. It handles the visual communication between the user of the application and the underneath logical components. Finally the controller, it is the liaison between the model and the view, the controller is the brain of the MVC pattern application, it processes the user inputs and gives instructions to the other logical components (model and view) on which operations to execute. Figure 7.A.1 represents a schematic representation of a MVC architecture. The use of a MVC architectural pattern has several benefits: it allows parallel development, it makes the application more robust to changes to the code, its flexibility allows to use multiple view logics for a same application.

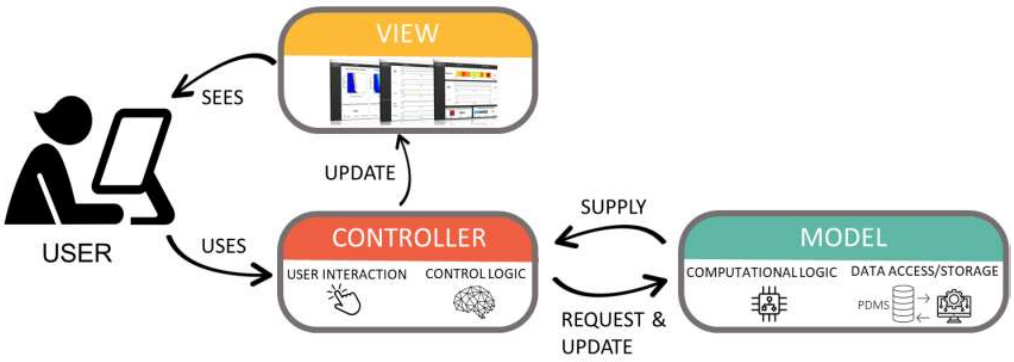


Figure 7.A.1 Example of MVC architecture. The user sees the interface of the application though the view logic and interacts with the application through the controller. The controller sends instructions to the model, which afterwards sends the results back to the controller. The controller might update the interface of the application by sending instructions to the view logic.

7.A.2 *Supplemental table*

Table 7.A.1 Report of software defects and corresponding solution

SOFTWARE DEFECT	DESCRIPTION	SUCCESSFUL SOLUTION	UPGRADED TB-AI VERSION
1	Software defect that occurs when the software is re-initialized.	YES. A new condition is implemented, if the study ID of the patient already exists in the database, the software does not create a new patient file.	1.1
2	PDMS exception. The software lost connection with the PDMS server	NO. The error is due to recurrent local updates of the PDMS software. When the updates are installed on the PDMS server the connection is interrupted. External problem related to the security protocol of the local PDMS, not solvable.	--
3	ICP and MAP were not correctly stored	YES. Software defect on data storage is solved.	1.2
4	Real-time ICP and LAx burden plot visualization error. Color scales are normalized, and therefore they change in relation to the maximum value to plot.	YES. Software defect on data visualization is solved. The range of plotted values is fixed across patients.	1.3
5	PDMS exception. The PDMS firewall blocks the software	YES. A new update of the PDMS firewall blocks the TB-AI software from connecting to the server. The problem was solved by changing the encryption used for the connection.	1.4
6	Software defects that occur when MAP is not measured	YES. A new condition is implemented in case the ICP signal is measured but the MAP is not (due to PMDS storage problems)	1.5

7.A.3 *Supplemental Figures*

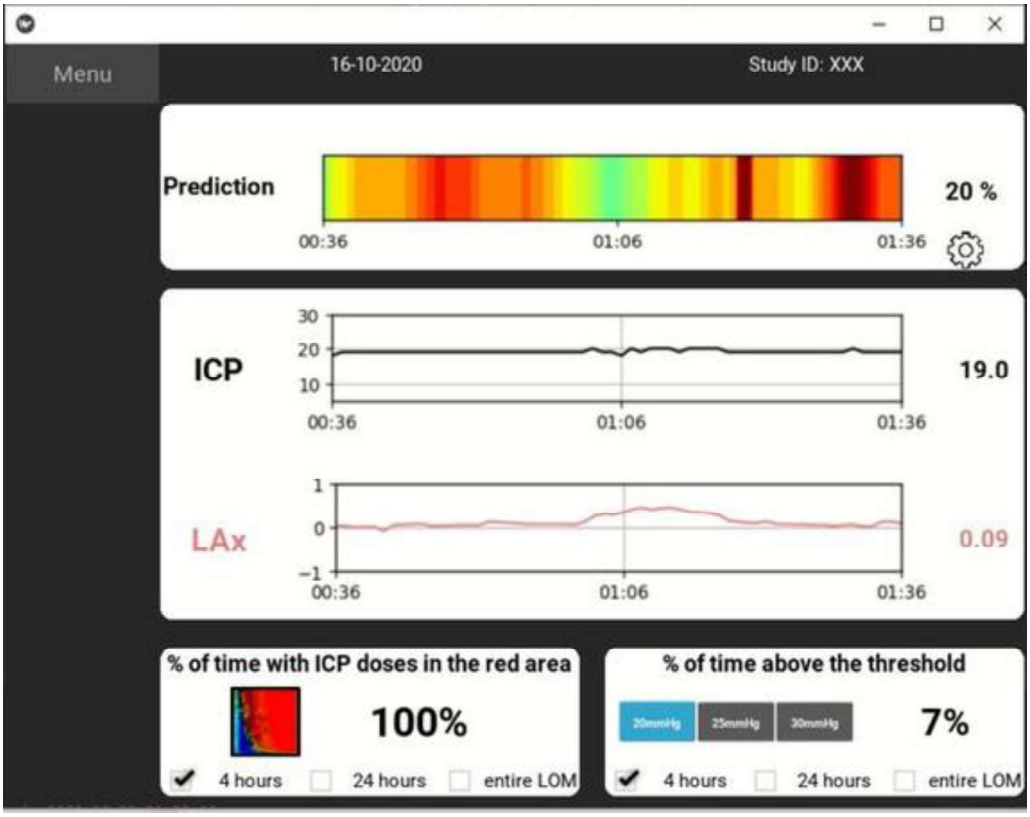


Figure 7.A.2 Home screen user interface of the TB-AI prototype software. Top panel: the prediction of upcoming events of elevated intracranial pressure (ICP) is displayed in a color-coded fashion. The settings of the color-coded visualization can be changed by the user by clicking on the setting wheel in the bottom right corner of the box. Central panel: the continuous minute-by-minute trace of ICP and low-frequency regulatory index (LAx, quantitative index of cerebrovascular autoregulation) are displayed. The LAx is computed every minute from the continuous ICP and mean arterial blood pressure signal (MAP). Bottom-left panel: percentage of time the patients spent in the red area. It is possible to select the window of time in which this metric is computed among the following options: last 4 hours of monitoring, last 24 hours of monitoring or entire length of monitoring (LOM). Bottom-right panel: displays the percentage of monitoring time the patient spent with an ICP above a certain threshold, which can be chosen by the user among the options of 20 mmHg, 25 mmHg or 30mmHg. This metric can be similarly computed for different windows of time, i.e. last 4 hours of monitoring time, last 24 hours of monitoring time or entire LOM. All metrics are computed in real time and updated each minute.

Figure 7.A.3 Monitoring signals screen user interface of the TB-AI software. The screen shows continuous minute-by-minute traces of relevant physiological signals for the management of patients with TBI. In detail, the screen shows: intracranial pressure (ICP), low-frequency regulatory index (LAx), mean arterial blood pressure (MAP), end-tidal CO2 and the partial brain tissue oxygen saturation ($PbtO_2$). The LAx signals is continuously calculated by the software, in fact this relevant continuous metric is not available yet at the bedside nor in the PDMS of UZ Leuven.

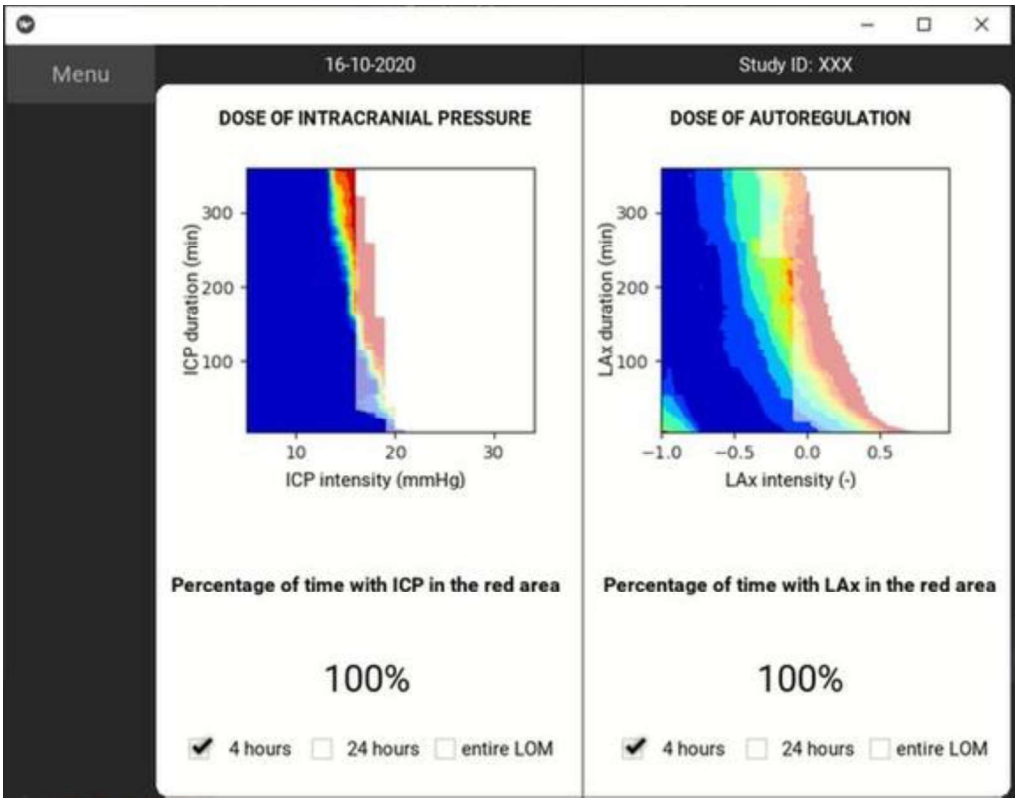


Figure 7.A.4 Intracranial pressure (ICP) and low-frequency autoregulatory index (LAx) burden visualization screen user interface of the TB-AI software. Left panel: real-time visualization graph of the ICP burden. The current ICP burden dose is plotted in vivid colors, where the cumulative ICP burden of the previous monitoring time, i.e. the worst ICP burden dose experienced by the patient, is plotted with more transparent colors. In the bottom left corner of the screen the percentage of time spent by the patient with an ICP in the "red area" of the visualization is displayed. The user can choose to have this percentage computed for different time windows, respectively for the past 4 hours, 24 hours or the entire length of monitoring (LOM). Right panel: real-time visualization of the LAx burden. The visualization follows the same principles of the ICP burden visualization plots explained before. Similarly to the left panel, the bottom right panel shows the percentage of time spent by the patient with a LAx in the "red area" of the visualization. As mentioned before, this percentage can be computed for different time windows.

BIBLIOGRAPHY

1. Maas AIR, Menon DK, Adelson PD, et al (2017) Traumatic brain injury: integrated approaches to improve prevention, clinical care, and research. Lancet Neurol 4422:. https://doi.org/10.1016/S1474-4422(17)30371-X

2. Stocchetti N, Maas AIR (2014) Traumatic Intracranial Hypertension. N Engl J Med 370:2121–2130. https://doi.org/10.1056/NEJMra1208708

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

4. Vik A, Nag T, Fredriksli O, et al (2008) Relationship of " dose " of intracranial hypertension to outcome in severe traumatic brain injury. J Neurosurg J Neurosurg J Neurosurg 109:0–678. https://doi.org/10.3171/JNS/2008/109/10/0678

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

6. Donnelly J, Güiza F, Depreitere B, et al (2021) Visualising the pressure-time burden of elevated intracranial pressure after severe traumatic brain injury: a retrospective confirmatory study. Br J Anaesth 126:e15–e17. https://doi.org/10.1016/j.bja.2020.09.018

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

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

9. Güiza F, Depreitere B, Piper I, et al (2013) Novel Methods to Predict Increased Intracranial Pressure During Intensive Care and Long-Term Neurologic Outcome After Traumatic Brain Injury: Development and Validation in a Multicenter Dataset. Crit Care Med 41:554–564. https://doi.org/DOI 10.1097/CCM.0b013e3182742d0a

10. Czosnyka M, Smielewski P, Kirkpatrick P, et al (1997) Continuous assessment of the cerebral vasomotor reactivity in head injury. Neurosurgery 41:11–7; discussion 17-9

11. Depreitere B, Güiza F, Van den Berghe G, et al (2014) Pressure autoregulation monitoring and cerebral perfusion pressure target recommendation in patients with severe traumatic brain injury based on

minute-by-minute monitoring data. J Neurosurg 120:1451–1457. https://doi.org/10.3171/2014.3.jns131500

12. Flechet M, Meyfroidt G, Piper I, et al (2018) Visualizing Cerebrovascular Autoregulation Insults and Their Association with Outcome in Adult and Paediatric Traumatic Brain Injury. In: Acta Neurochirurgica Supplementum. pp 291–295

8

REAL-TIME PERFORMANCE OF A PREDICTION MODEL FOR HARMFUL DOSES OF INTRACRANIAL PRESSURE: A DESCRIPTIVE ANALYSIS

ADAPTED FROM: **Carra G.**, et al. “Real-time performance of a prediction model for harmful doses of intracranial pressure: a descriptive analysis”. Manuscript ready for submission.

ABSTRACT

OBJECTIVE: To assess the performance of a prediction model for harmful doses of intracranial pressure (ICP) in patients with severe traumatic brain injury (TBI) in a real-time clinical setting.

APPROACH: Patients with severe TBI and invasive ICP monitoring were included in a single-center prospective observational study. Minute-by-minute ICP and mean arterial blood pressure signals were collected in addition to clinical and demographic data. Afterwards, we computed the continuous predictions of the model and assessed the model performance with respect to area under the receiver operating characteristic (AUC), area under the precision-recall curve (AP), accuracy, precision, sensitivity and specificity. In addition, we assessed the clinical usefulness of the model with decision curve analysis. Last, we computed the level of technology readiness (TRL) for bedside implementation.

MAIN RESULTS: Fourteen patients with severe TBI were included in this prospective study, with a continuous monitoring time of 5 [IQR 3-10] days. The model presented an AUC of 0.92, AP of 0.76, accuracy of 0.87, precision of 0.68, sensitivity of 0.69 and specificity of 0.91 (alerting threshold= 0.50). The model presented calibration slope of 0.69 and a calibration-in-the-large of 0.05 (p-value < 0.01). Clinical usefulness was demonstrated for the risk thresholds [0.09-0.65]. With this study, the model achieved a TRL = 6.

SIGNIFICANCE: In a clinical setting with real-time continuous data, the model demonstrated excellent discrimination and specificity, good calibration and clinical usefulness for a broad range of risk thresholds. In a real-time clinical setting, the high specificity of the model, at a cost of a lower sensitivity, is to be preferred in light of potential alert fatigue. The results indicate that the model has a sufficiently high TRL for a future interventional study to assess the impact of its use at the bedside on patient management, and on short and long-term outcomes.

Abnormally elevated intracranial pressure (ICP) occurs in approximately 60% of patients with severe traumatic brain injury (TBI) [1, 2], and it is associated with increased mortality and worse neurological outcomes [3].

Treatment and prevention of elevated ICP are key aspects in the management of TBI. Current guidelines recommend the use of aggressive treatments when the ICP crosses the fixed threshold of 22 mmHg [4]. However, brief ICP elevations above a fixed threshold might not always result in brain damage and the ability of the patient to tolerate intracranial hypertension may vary according to age, sex or cerebrovascular autoregulation status, among others. As such, a universal ICP threshold, which applies to all patients, in every situation, is too simplistic, especially considering that this threshold was derived from a single center epidemiological study that found the highest association with outcome at 22 mmHg [5]. An epidemiological association should not be confounded with a treatment threshold.

Several studies have demonstrated that the duration of these events plays a crucial role in the definition of harmful ICP events [6, 7]. The concept of "dose of ICP" combines intensity and duration of an event of elevated ICP [6], and as such it integrates a second dimension, that is time, in the evaluation of the danger of elevated ICP. Elevated doses of ICP have been associated with worse neurological outcomes and increased mortality [7–11] and in one single-center study the ICP dose presented a stronger association with outcomes than time-point ICP values above a fixed threshold [9]. Güiza et al. [7] displayed the association between intensity and duration of an ICP event and long-term outcomes in a color-coded plot. In the plot, ICP doses that are represented in blue occur more frequently in patients with better long-term outcomes, while ICP doses that are represented in red occur more frequently in patients with worse long-term outcomes. The blue and red regions are clearly divided by an exponential line. Despite having demonstrated clinical value in several studies [6–11], the concept of ICP dose is not currently used in clinical practice and currently only one monitor (only recently) available in the market displays information about the ICP dose, although in a limited form.

However, it is challenging to identify the early signs of an ICP event whose combination of intensity and duration may be harmful for the patient. For this reason, in a previous study we developed a machine learning (ML)

model for the early prediction of events of ICP whose doses have been associated with poor neurological outcomes in several epidemiological studies (see Chapter 6). The model processes minute-by-minute ICP and mean arterial blood pressure (MAP) to provide with 30 minutes forewarning alerts on the future occurrence of ICP doses belonging to the red area of the visualization of Güiza et al. [7]. The model was developed on a large multi-center dataset [12] and demonstrated good performance when validated on the Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI) dataset [13], a large, European, prospectively collected, multicenter dataset of patients with TBI. On the CENTER-TBI the model had AUC = 0.94, AP = 0.89, accuracy = 0.89, precision = 0.79, sensitivity = 0.77 and specificity = 0.93 (alerting threshold = 0.5). Although this model presents good performance when evaluated on the development and external validation dataset, there is no evidence that its use at the bedside would improve clinical practice and potentially affect outcomes.

The prospective evaluation of the model performance on a real clinical setting is an important step for the future integration of the model at the bedside and the evaluation of its impact on patients' management and outcomes. This step is important to improve the level of technology readiness (TRL) of the model [14], but challenging because of technical and regulatory reasons. Recently, Fleuren et al. [14] proposed a grading scale for clinical readiness of ML models. The scale ranges from a minimum of TRL = 1 (identification of the clinical problem), to a maximum of TRL = 9 (model integration in the clinical workflow as safe and accurate medical device). The authors found that only 2% of the 160 reviewed articles scored level 6 or above [14]. Similarly, a recent systematic review from van de Sande et al. [15] showed that of 441 studies involving the development of ML models on ICU data only 8 (1.6%) prospectively tested the model in a clinical setting.

In this descriptive study, we set up a prospective, observational, single-center study to obtain an assessment of the real-time performance of the model for the prediction of harmful ICP doses. Moreover, we will assess the TRL of the model for what concerns clinical implementation.

8.2 METHODS

8.2.1 *Study population*

This prospective observational study included patients with severe TBI that were admitted in the intensive care unit (ICU) of the University Hospitals UZ Leuven (UZ Leuven), Belgium, between January 2020 and April 2021. Given the exploratory nature of the study, we did not perform a sample size calculation, but we aimed at enrolling at least 10 patients. Patients were declared eligible for inclusion if they underwent invasive intra-parenchymal ICP monitoring and if they were not subjected to a therapy restriction code at the moment of inclusion. The medical ethics committee of the University Hospitals UZ Leuven, Belgium approved this study. Use of informed consent was waived, but the patients and/or their families were informed about the prospective collection of data though written communication.

8.2.2 *Prospective study*

During the prospective study the following demographic and clinical data were collected: age, sex, Marshall score at admission, minute-by-minute ICP and MAP signals, worse daily Treatment Intensity Level [16] (TIL, which ranges from a minimum of 1 to a maximum of 38), Glasgow Coma Scale (GCS) at the start and the end of the ICP monitoring.

Minute-by-minute ICP and MAP signals where prospectively collected (minute-by-minute collection and storage) by a prototype bedside software that queried data from the local Patient Data Management System (PDMS, MetaVision®; iMD-Soft®, Needham, MA, USA) of UZ Leuven. The blinded software was initiated after start of ICP monitoring and it was stopped after ICP monitoring completion. The use of the bedside software for the collection of data allows to obtain minute-by-minute signals collected in a real clinical setting and therefore mimic the use of the model at the bedside. Continuous predictions were computed from the minute-by-minute ICP and MAP signals that were prospectively collected at the bedside.

8.2.3 *Model external validation and assessment of the real-time performance*

The “real-time performance” of the model was computed with respect to the continuous (minute-by-minute) predictions, see Figure 8.1 for an example. Additional examples of continuous predictions can be found in Appendix

8.A.1. The evaluation of the performance of the model on continuous ICP sequences provides a better estimate of the error of the model when used in clinical practice.

Model performance was assessed according to the original study (Chapter 6) and computed for the entire population. In details, we computed the area under the receiving characteristic curve (AUC), area under the precision recall curve (AP), accuracy, precision, sensitivity, and specificity. Calibration was assessed by plotting calibration plots, and by computing the calibration slope and calibration-in-the-large, while clinical importance was assessed with decision curves. Decision curve analysis compares the clinical usefulness of the prediction as an alerting tool with the clinical usefulness of the default strategies of "alert for all patients" or "alert for none". Clinical usefulness is assessed by the computation of the net benefit.

To investigate the changes in performance as a result of different alerting thresholds, threshold-dependent performance metrics, i.e. accuracy, precision, sensitivity, and specificity, were computed for different scenarios. In particular: *a*) we used an alerting threshold of 0.5 in accordance with the study presented in Chapter 6; *b*) we used as alerting threshold the optimal threshold as identified from the receiving characteristic curve;

The TRL of the model for clinical implementation was assessed according to the grading scale proposed by Fleuren et al. [14].

Statistical analysis was performed in Python (*version 3.5,* https://www.python.org) with the following libraries: numpy (*version 1.15,* https://numpy.org/) and scipy (*version 0.20*, https://www.scipy.org/). Calibration curves were generated with the R-based library givitiR (*version* 1.3, https://CRAN.R-project.org/package=givitiR).

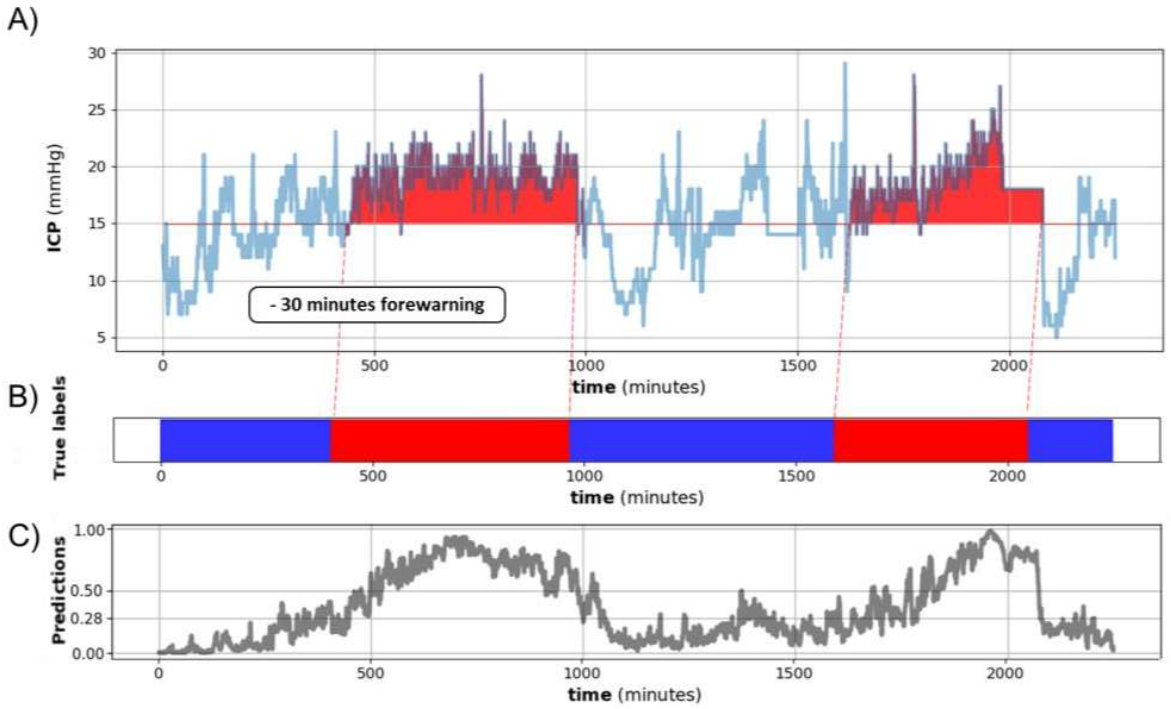


Figure 8.1 Example of the real-time predictions of the model as compared to the true labels. Panel A) continuous ICP sequence, events of ICP whose dose belongs to the red area of the visualization of Güiza et al. [7] are indicated with the red shaded area. Panel B) true labels for the ICP sequence in exam. The time points in blue and red indicate, 30 minutes in advance, the presence of an ICP event whose dose belongs to the blue and red area of the visualization of Güiza et al. [7] respectively. Panel C) displays the real-time predicted probabilities of the model for events of ICP dose in the red area. The model may fire an alarm depending on the chosen alerting threshold.

8.3 RESULTS

8.3.1 *ICP characterization*

Fourteen patients with severe TBI were included in the study. Patients had an age at admission of 55 years [IQR 32-58], a Marshall score at ICU admission of 2 [IQR 2-4], a GCS at study admission of 7 [IQR 5-9] and a GCS at study end of 9 [IQR 3-10]. Eighty-five percent of patients were male and their average daily TIL was 7.5 [IQR 6.0 – 9.0]. Patients experienced a percentage of monitoring time in the red area of 15% [IQR 0-42%] for a total of 5 [IQR 3-10] monitoring days. Three patients did not experience events of ICP in the red area. Patients' summary statistics are summarized in Table 8.1.

Table 8.1 Patients' summary statistics

VARIABLE	VALUE
Age, years, median [IQR]	55 [32-58]
Sex, males, percentage	85
Marshall score, median [IQR]	2 [2-4]
GCS at study admission, median [IQR]	7 [5-9]
GCS at study end, median [IQR]	9 [3-10]
Daily TIL, median [IQR]	7.5 [6.0-9.0]
Monitoring time with ICP in the red area, percentage	15 [0-42]
Monitoring days, median [IQR]	5 [3-10]

GCS: Glasgow Coma Scale

TIL: treatment intensity level

ICP: intracranial pressure

8.3.2 *Model external validation and assessment of the real-time performance*

The dataset for the evaluation of the real-time performance of the model comprised 144019 samples of minute-by-minute ICP and MAP (about 100 days of continuous predictions), of which 20551 (14.27%) corresponded to events of ICP dose in the red area.

When evaluated on this prospectively collected dataset of real-time monitoring signals the model had an AUC of 0.92 and an AP of 0.76. With an alerting threshold equal to 0.5, the model had an accuracy of 0.87, precision of 0.68, sensitivity of 0.69 and specificity of 0.91. The optimal alerting threshold as computed from the receiving characteristic curve was equal to 0.28. At this alerting threshold, the model had an accuracy of 0.82, precision of 0.55, sensitivity of 0.87 and specificity of 0.81. See Table 8.2 for a summary of the model performance for different alerting thresholds. Calibration analysis resulted in a calibration slope of 0.69 and a calibration-in-the-large of 0.05, p-value < 0.01, see Figure 8.2 panel A. The model presented clinical benefit as compared to the "alert for all" and "alert for none" options in the risk range [0.09-0.65], see Figure 8.2, panel B.

According to the grading scale for clinical readiness of Fleuren et al. [14], the model has a TRL = 6. A TRL = 6 requires the achievement of the following methodological milestone: "*Model performance is tested real-time and integrated into the EHR/hospital system in one or more clinical settings, but no implementation into clinical workflow (i.e. no clinical staff is exposed to model results)*" [14]. According to the grading scale of Fleuren et al. [14] a TRL = 6 is obtained by evaluating the model in a prospective observational study and demonstrating that the designed technical pipeline for the automated extraction of data in a clinical context is performed successfully.

Table 8.2 Model performance. Performance metrics are computed for the entire population

PERFORMANCE METRIC	VALUE
AUC	0.92
AP	0.76
alerting threshold = 0.50	
Accuracy	0.87
Precision	0.68
Sensitivity	0.69
Specificity	0.91
alerting threshold = 0.28	
Accuracy	0.82
Precision	0.55
Sensitivity	0.87
Specificity	0.81

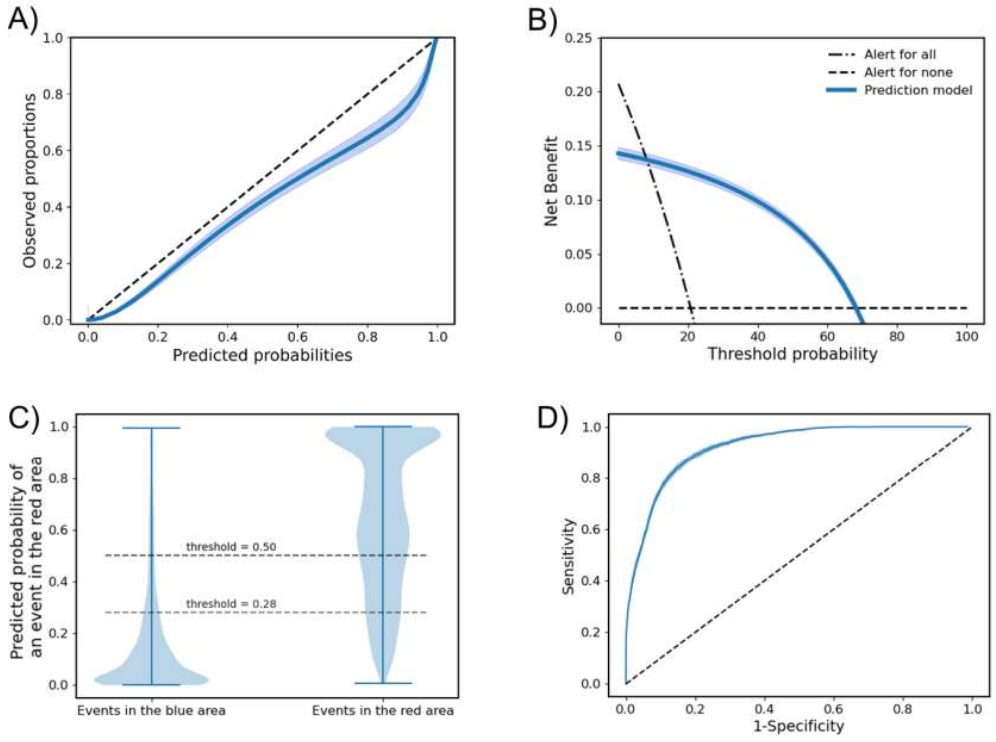


Figure 8.2 Performance of the prediction model. Panel A) Calibration curve. Panel B) Decision curve. Clinical benefit is demonstrated between the risk thresholds range of [0.09 - 0.65]. Panel C) Violin plots of the predicted probability

scores for events in the blue area and red area of the visualization plot. Two alerting thresholds are displayed. An alerting threshold equal to 0.5 corresponds to the alert threshold used in the external validation dataset (CENTER-TBI) while an alerting threshold of 0.28 corresponds to the threshold that maximizes sensitivity and specificity. It is possible to appreciate the change in number of false negatives and false positives in relation to the chosen alerting threshold. Panel D) Receiving characteristic curve of the model.

8.4 DISCUSSION

The goal of this prospective observational study was the evaluation of the real-time performance of the ML model for the prediction of harmful doses of ICP on a real clinical setting. This is the first study that assesses the bedside performance of a prediction model for harmful ICP.

When evaluated on this dataset, which approximately counted 100 continuous days of minute-by-minute predictions, the model presented slightly lower AUC (0.92), AP (0.76), accuracy (0.87), precision (0.68), sensitivity (0.69) and specificity (0.91) as compared to the performance on the external validation dataset consisting of the CENTER-TBI database (see Chapter 6). A drop in performance when evaluating the model on continuous data is expected. In fact, during the model development the events of ICP are strictly classified in blue or red regions, as visualized in Figure 8.1 panel B. This is a methodological simplification that does not fully represent a real clinical situation, were the transition between the two regions is likely to be more gradual. Also, in real-time clinical data the prevalence of events in the blue versus red area is more unbalanced as compared to the development and external validation dataset, which can also affect performance. Moreover, it is possible to appreciate from the violin plots of Figure 8.2 panel C that the majority of the predictions cluster around either 0 or 1 which corresponds to high certainty in the prediction and consequently low misclassification for a broad range of alerting thresholds. Similarly, the decision curve showed clinical benefit for the risk thresholds [0.09-0.65], indicating that when tested on continuous signals the model presents useful clinical information for a very broad range of risk thresholds. Finally, for this specific clinical problem high specificity is preferred to high sensitivity, to strictly avoid alarm-fatigue, while still providing useful information to the clinicians. Therefore, we believe that

the model presents good enough performance to justify proceeding to an interventional clinical study.

Importantly, the model performance highly depended on the chosen alerting threshold, with the most important impact on sensitivity and specificity. Acceptable alerting thresholds will need to be evaluated carefully by the clinician depending on the local policies and the clinical and personal needs of the patient, and it will mostly depend on the intended use of the model. Such threshold should be chosen from the clinical useful range [0.09-0.65], where for instance a threshold = 0.09 favors high sensitivity, while a threshold = 0.65 threshold will be chosen in situations where high precision and high specificity are desired (at the cost of a higher number of false negatives).

A potential limitation of the presented study is the limited number of included patients. However, since every patients was monitored for a minimum of 3 days, the performance metrics were computed on more than 144000 minute-by-minute predictions.

This study represents a fundamental step in the process of bringing the prediction model for harmful ICP doses to the bedside. Importantly, this study increases the TRL of the model to level 6. The next step for the translation from bench to bedside is the implementation of the model in the clinical workflow through a future not-blinded prospective study. A future interventional study will need to evaluate whether the use at the bedside of the prediction model as an alerting tool will improve short- and long-term outcomes of patients as compared to an historical control group such as the CENTER-TBI. Outcomes of interest could be ICU mortality and length of stay, number of deleterious events of elevated ICP, time in the safe "blue area" of the visualization plot and GOS at 6-months.

8.5 CONCLUSIONS

In this study, we evaluated the real-time performance of a prediction model for harmful ICP doses. When evaluated on prospectively collected minute-by-minute data the model demonstrated excellent specificity, good calibration and clinical usefulness for a broad range of risk thresholds. These findings set the basis for a future clinical implementation and the

prospective evaluation of the benefits of the model on quality of care and patient outcomes.

ACKNOWLEDGMENT AND PERSONAL CONTRIBUTION

Study concept and design	Meyfroidt, Güiza, **Carra,** Depreitere
Data acquisition	Mebis, Wouters, **Carra,** Dietvorst, Chao-Yuan Huang, Vanhulle, Hendrickx, Witpas
Statistical analysis	**Carra**
Interpretation of results	**Carra,** Güiza, Meyfroidt
Drafting of the manuscript	**Carra**
Manuscript revision	**All authors**
Principal investigator	Meyfroidt

8.A APPENDIX

8.A.1 *Supplementary figures*

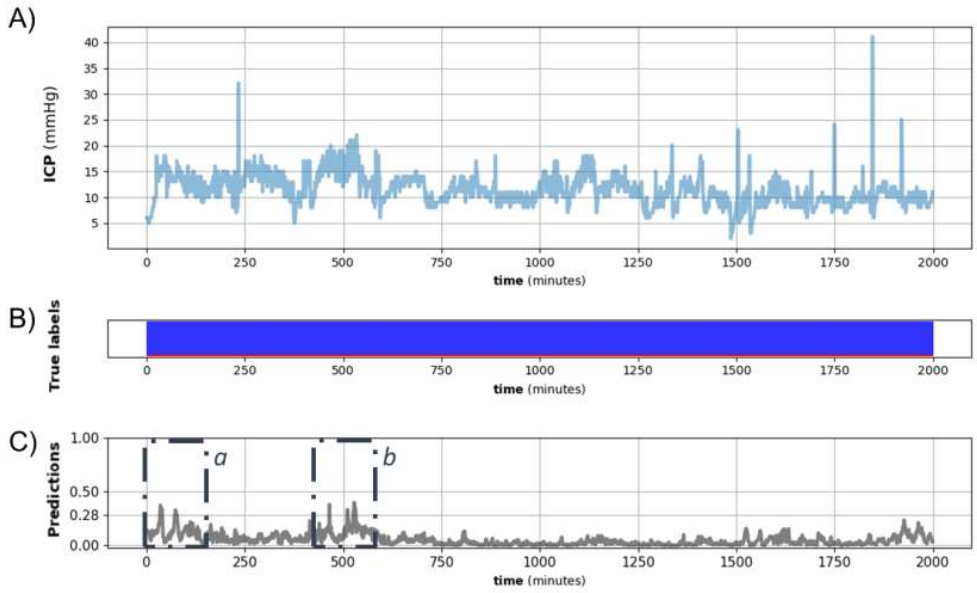


Figure 8.A.1 Example of real-time predictions of the model as compared to the true labels. Panel A) continuous ICP sequence, in this example the patient does not experience events of ICP dose in the red area of the visualization of Güiza et al. [7]. Panel B) true labels for the ICP sequence in exam. The time points in blue indicate, 30 minutes in advance, the presence of an ICP event whose dose belongs to the blue area of the visualization of Güiza et al. [7]. Panel C) displays the real-time predicted probabilities of the model for events of ICP dose in the red area. The model may fire an alarm depending on the chosen alerting threshold. Overall, the model correctly provides low probabilities of future events of ICP in the red area. Depending on the alerting threshold, brief episodes of false positive predictions may be provided by the model in sequence *a* and *b* of panel C.

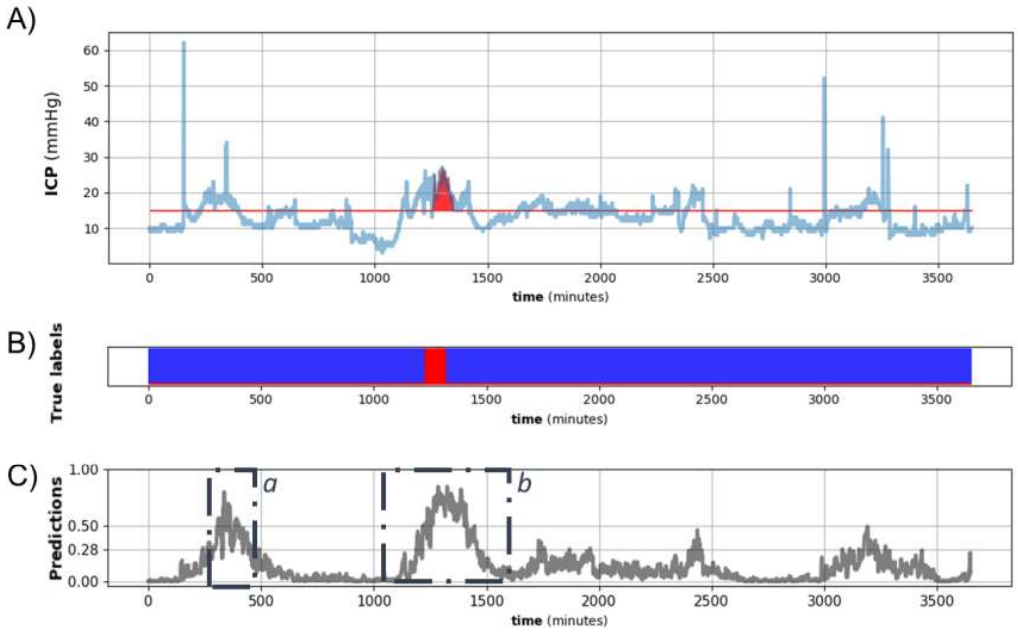


Figure 8.A.2 Example of real-time predictions of the model as compared to the true labels. Panel A) continuous ICP sequence, events of ICP whose dose belongs to the red area of the visualization of Güiza et al. [7] are indicated with the red shaded area. Panel B) true labels for the ICP sequence in exam. The time points in blue and red indicate, 30 minutes in advance, the presence of an ICP event whose dose belongs to the blue and red area of the visualization of Güiza et al. [7] respectively. Panel C) displays the real-time predicted probabilities of the model for events of ICP dose in the red area. The model may fire an alarm depending on the chosen alerting threshold. In this example, it is likely that the model will provide a series of false positives predictions in sequence *a* of panel C. Moreover, in sequence *b* of panel C the model correctly provides high probabilities of a future event in the red area, depending on the chosen alerting threshold the alert will be provided with forewarning equal or longer than 30 minutes. Overall, the model correctly provides probabilities of future events of ICP in the red area that match the true labels.

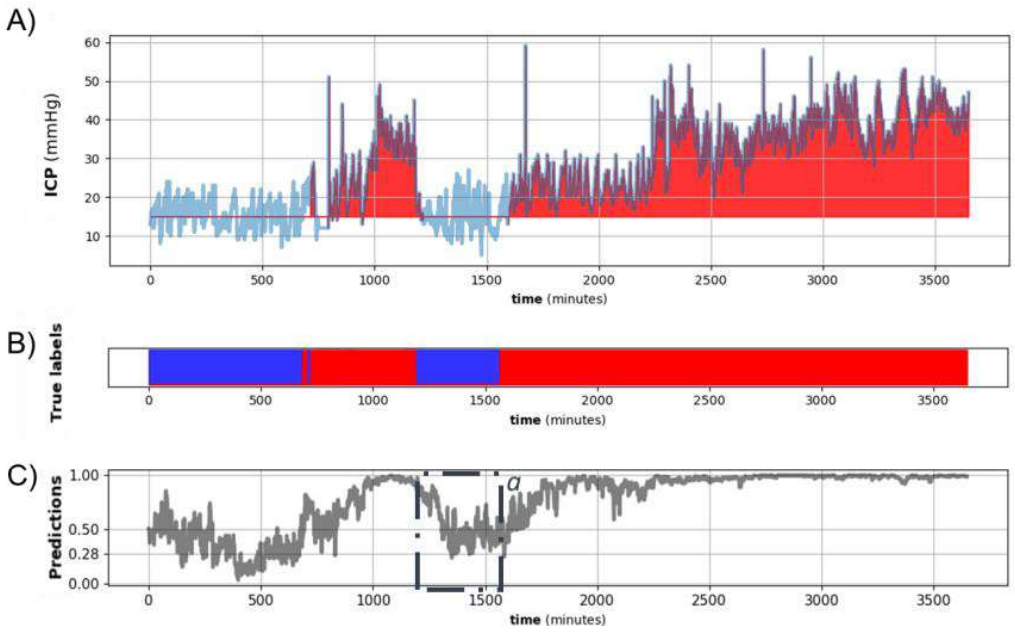


Figure 8.A.3 Example of real-time predictions of the model as compared to the true labels. Panel A) continuous ICP sequence, events of ICP whose dose belongs to the red area of the visualization of Güiza et al. [7] are indicated with the red shaded area. Panel B) true labels for the ICP sequence in exam. The time points in blue and red indicate, 30 minutes in advance, the presence of an ICP event whose dose belongs to the blue and red area of the visualization of Güiza et al. [7] respectively. Panel C) displays the real-time predicted probabilities of the model for events of ICP dose in the red area. The model may fire an alarm depending on the chosen alerting threshold. In this example, depending on the chosen alerting threshold, the model may provide a series of false positives predictions in sequence *a* of panel C. Overall, the model correctly provides high probabilities for future events of ICP in the red area.

BIBLIOGRAPHY

1. Stocchetti N, Colombo A, Ortolano F, et al (2007) Time Course of Intracranial Hypertension after Traumatic Brain Injury. J Neurotrauma 24:1339–1346. https://doi.org/10.1089/neu.2007.0300

2. Andrews PJD, Sinclair HL, Rodriguez A, et al (2015) Hypothermia for Intracranial Hypertension after Traumatic Brain Injury. N Engl J Med 373:2403–2412. https://doi.org/10.1056/NEJMoa1507581

3. Maas AIR, Menon DK, Adelson PD, et al (2017) Traumatic brain injury: integrated approaches to improve prevention, clinical care, and research. Lancet Neurol 4422:. https://doi.org/10.1016/S1474-4422(17)30371-X

4. Carney N, Totten AM, O'Reilly C, et al (2016) Guidelines for the Management of Severe Traumatic Brain Injury 4th Edition

5. Sorrentino E, Diedler J, Kasprowicz M, et al (2012) Critical Thresholds for Cerebrovascular Reactivity After Traumatic Brain Injury. Neurocrit Care 16:258–266. https://doi.org/10.1007/s12028-011-9630-8

6. Vik A, Nag T, Fredriksli O, et al (2008) Relationship of " dose " of intracranial hypertension to outcome in severe traumatic brain injury. J Neurosurg J Neurosurg J Neurosurg 109:0–678. https://doi.org/10.3171/JNS/2008/109/10/0678

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

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

9. Sheth KN, Stein DM, Aarabi B, et al (2013) Intracranial Pressure Dose and Outcome in Traumatic Brain Injury. Neurocrit Care 18:26–32. https://doi.org/10.1007/s12028-012-9780-3

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

11. Lazaridis C, DeSantis SM, Smielewski P, et al (2014) Patient-specific thresholds of intracranial pressure in severe traumatic brain injury. J Neurosurg 120:893–900. https://doi.org/10.3171/2014.1.JNS131292

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

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

14. Fleuren LM, Thoral P, Shillan D, et al (2020) Machine learning in intensive care medicine: ready for take-off? Intensive Care Med 46:1486–1488. https://doi.org/10.1007/s00134-020-06045-y

15. van de Sande D, van Genderen ME, Huiskens J, et al (2021) Moving from bytes to bedside: a systematic review on the use of artificial intelligence in the intensive care unit. Intensive Care Med 47:750–760. https://doi.org/10.1007/s00134-021-06446-7

16. Maas AIR, Harrison-Felix CL, Menon D, et al (2011) Standardizing Data Collection in Traumatic Brain Injury. J Neurotrauma 28:177–187. https://doi.org/10.1089/neu.2010.1617

9

GENERAL DISCUSSION

After critical illness, patients may suffer from long-term neurological impairment, with potentially significant consequences on their quality of life. During critical care, it is therefore of great interest to detect early signs of impending neurological damage, upon which clinicians can react and potentially impact neurocognitive outcomes. This interest goes beyond the prevention of mortality, but rather aims at improving the patient's quality of life after ICU discharge.

In a modern ICU, the brain can be monitored using several modalities, invasively and non-invasively. The data that originates from these monitors contains useful information, which can be used to optimize treatment and improve the long-term outcomes of the patients. However, the interpretation and analysis of these multi-dimensional data may be challenging, and a visual examination or the analysis of simple descriptive statistics may not suffice. More advanced techniques may be needed. Artificial intelligence (AI) could be used for this purpose, as these techniques can be used to detect hidden patterns that may be linked with occurring neurological damage but are also able to provide predictions of impending brain-threatening events.

In this thesis, we used AI to provide insights for decision-support and treatment optimization in three different types of patients with, or at risk of, brain injuries. In the first objective, we assessed the independent clinical usefulness of near-infrared spectroscopy to predict long-term neurological damage in children with congenital heart defects. In the second objective, we investigated the association between ICP doses and outcome in patients with aSAH. In the third objective, we validated a decision-support model for the early-prediction of elevated ICP in patients with TBI. In the fourth objective, we developed a model for the early prediction of a broad range of harmful doses of ICP in patients with TBI. In the fifth objective, we developed a prototype software for decision support and treatment

optimization in patients with TBI, implemented this software at the bedside, and tested the technical functioning as well as the predictive performance of the software.

9.1 ASSESSMENT OF THE CLINICAL USEFULNESS OF NEAR-INFRARED SPECTROSCOPY IN THE POST-OPERATIVE CARE OF CHILDREN WITH CONGENITAL HEART DISEASES

9.1.1 *Main findings*

Children with severe congenital heart defects often suffer from neurodevelopmental deficits, because of associated genetic or congenital neurological or metabolic defects, or problems during gestation. Many of these congenital heart defects require surgical correction towards a normal or palliative physiology. The peri-operative period is often complicated with additional hypoxic or hypo-perfusion events, which may have an additional independent impact on outcomes. Peri-operative care for this vulnerable population is complex and pursuing adequate cerebral perfusion during the entire perioperative period is a challenge. Invasive monitoring in a pediatric setting is not self-evident. Non-invasive cerebral tissue oxygen saturation ($SctO_2$) monitoring with near-infrared spectroscopy (NIRS) is a frequently used technique. Reduced perioperative $SctO_2$ has been associated with worse short-term outcomes, such as increased length of PICU stay, post-operative complications and mortality [1–3]. Nonetheless, critical $SctO_2$ thresholds to guide medical interventions are difficult to define. As a result, recommendations on treatment optimization based on NIRS monitoring are non-specific. It remains unclear whether NIRS provides clinically useful information, and how to use this information to deliver better treatment. For this scope, it is of great interest to investigate the association between reduced $SctO_2$ and long-term neurodevelopmental outcomes, which remains undefined.

In Chapter 3 we investigated the association between $SctO_2$ as monitored by the NIRS and total IQ, 2 years after PICU admission. For this study, we set up a prospective observational blinded study in the PICU of the University Hospitals Leuven (UZ Leuven) between 2012 and 2015. The study included children after surgery for congenital heart defects. Children were monitored with the FORESIGHT NIRS monitor (CAS Medical Systems, Branford, CT.) up until the interruption of mechanical ventilation. The parents or legal guardian of the children were contacted 2 years after

PICU admission for a follow-up visit, where the total IQ was assessed by a team of trained psychologists. In the retrospective analysis of data, two features were extracted from the post-operative $SctO_2$ signals, namely the mean $SctO_2$ and the dose of desaturation (defined as $SctO_2 < 65\%$). The independent association between $SctO_2$ features and total IQ was investigated with a Bayesian linear regression model corrected for the age, nutrition strategy, severity ad admission, presence of syndrome, nutritional strategy, and presence of cyanosis after surgery. We found that increased $SctO_2$ desaturation and decreased $SctO_2$ mean in the early-postoperative period independently increased the probability of a lower total IQ at 2-years follow-up. These findings were confirmed by several sensitivity analyses for different definitions of desaturation and for additional correcting factors in the multivariable regression model.

9.1.2 *Current impact of research and future perspectives*

The study presented in Chapter 3 demonstrates that events of reduced $SctO_2$ in the postoperative period after cardiac surgery for CHD have a measurable impact on the total IQ up to 2 years later. These results suggest that $SctO_2$ desaturations are important clinical events and that the NIRS provides clinically relevant information that could be used to optimize treatment. In particular, optimal $SctO_2$ values are in a rather tight range between 65% and the upper range of 85%-90%, where both reduced and elevated $SctO_2$ could be a marker of ongoing neurological damage. The results of this study may further corroborate the susceptibility of the immature brain to even brief episodes of reduced $SctO_2$. Whether strategies aimed at preventing or treating these desaturations could influence the negative impact on IQ remains to be proven. Importantly, this study shows associations and no casual relations. A randomized interventional study would be needed to fully assess the impact of $SctO_2$-oriented management strategies on outcomes. In this context, it is important to investigate whether it is realistic to design a management protocol to control or correct $SctO_2$ by medical interventions. Preliminary prospective studies have already suggested that such $SctO_2$ manipulation is possible, with an effect of outcomes [4, 5]. As these were single-center studies that only focused on the prevention of severe desaturation episodes ($SctO_2<50\%$ and $SctO_2<20\%$), larger multicenter, randomized prospective studies whose management protocol targets $SctO_2$ levels above 65% are needed.

To assess the clinical value of $SctO_2$, relevant $SctO_2$ summary metrics, such as the $SctO_2$ dose or $SctO_2$ mean over past intervals of time, need to be available at the bedside. Current NIRS monitors display only the current $SctO_2$ value, sometimes with a timeline of previous values, but with none or very little summary statistics. The desaturation dose below user-specific threshold, which resulted highly significant in our study, is currently only available in one cerebral oximeter on the market (Root with O_3 Regional Oximetry, Masimo, CA, USA). Our research, as well as several previous studies [1, 3], show the importance of interpreting the $SctO_2$ trace with more complex statistical methods.

Moreover, it would be interesting to assess whether $SctO_2$ desaturation and reduced $SctO_2$ are also associated with total IQ at later time points than 2 years follow-up (for example 5- or 10-years follow-up). This information would be important to assess the long-term burden of impaired cerebral perfusion when experienced at a young age.

In this study, we exclusively focused on the lower range of $SctO_2$. However, the effect of hyperoxia ($SctO_2$ > 85-90%) on the developing brain is equally important. Oxidative stress is known to be associated with pediatric conditions such as periventricular leukomalacia, peri- and/or intra-ventricular hemorrhage or severe retinopathy [6–9]. The association between hyperoxia and long-term outcomes could not be investigated in the context of the study presented in Chapter 3, given that it was not part of the pre-planned statistical analysis and that few patients experienced such events, and therefore statistical power could not be reached. However, the long-term consequences of abnormally elevated $SctO_2$ values on outcomes should be investigated in future studies to further assess the clinical usefulness of NIRS monitoring.

9.2 ASSESSMENT OF THE ASSOCIATION BETWEEN ICP DOSES AND LONG-TERM OUTCOMES IN PATIENTS WITH ASAH

9.2.1 *Main findings*

Patients with aneurysmal subarachnoid hemorrhage (aSAH) may suffer for increased intracranial pressure (ICP), which is associated with increased mortality and poor long term outcomes [10–12]. Specific recommendations for the treatment of elevated ICP in patients with aSAH are missing, and as a result, ICP is not monitored in every center. In addition, local protocols for the management of elevated ICP are mostly based on the guidelines for

traumatic brain injury, while safe ICP levels for patients with aSAH remain ill defined. In patients with aSAH treatment is based on a threshold-based approach that suggests the initiation of aggressive treatments when the ICP rises above 20-22 mmHg, similar to TBI. More insights on the role of elevated ICP in aSAH may provide a basis for ICP treatment optimization, with a potential impact on long-term outcomes.

In Chapter 4, we investigated the association between ICP dose and long-term neurological outcomes of patients with aSAH with the methodology introduced by Güiza et al. [10]. Such methodology allows the extraction of a color-coded plot that displays which events of ICP dose occur more frequently in patients with better and worse long-term outcomes. To achieve this goal, we performed a retrospective analysis of multicenter, prospectively collected data of 98 adult patients with aSAH amendable to treatment. Patients were admitted to the ICUs of two European centers (Medical University of Innsbruck [Austria] and San Gerardo University Hospital of Monza [Italy]) from 2009 to 2013. In addition, the independent association between the cumulative dose of intracranial hypertension and outcome for each patient was investigated by using multivariable logistic regression models corrected for age, occurrence of delayed cerebral ischemia (DCI), and the Glasgow Coma Scale score at admission. In both cohorts, the combination of duration and intensity defined the tolerance to intracranial hypertension. A semi-exponential transition curve divided ICP doses that were associated with better outcomes (in blue) with ICP doses associated with worse outcomes (in red). For the two cohorts, the transition curve occurred at different ICP levels. An increased similarity between transition curves was found when patients with known secondary complications were excluded from the visualization analysis. Importantly, an independent association was found between the cumulative time that the patients experienced ICP doses in the red area and long-term neurological outcomes in both cohorts. The ICP dose was an independent predictor of long-term outcomes, whereas the cumulative time spent by the patients with an ICP greater than 20 mmHg was not.

9.2.2 Current impact of research and future perspectives

The results presented in Chapter 4 suggest that the duration of an ICP elevation is a relevant measure for the evaluation of ICP harmfulness and

therefore should be considered in the decision-making process for treatment of elevated ICP. Interestingly, for both aSAH cohorts the transition curves occurred at lower ICP levels than the ones obtained for TBI [13–15]. Once again, this suggests that treatment of ICP in aSAH requires a targeted and personalized approach, with two possible interpretations. First, the visualization suggests that in patients with aSAH, ICP levels around 15-20 mmHg may already be harmful and that clinicians should target as low ICP values as possible. In alternative, the results may indicate that intracranial hypertension is an important surrogate marker of underlying mechanisms, that eventually lead to neurological deterioration. As an argument in this favor, it is likely that cerebrospinal fluid (CSF) circulation disturbances were not the only cause of increasing ICP values, given that in both cohorts continuous CSF drainage did not suffice to control the ICP to more physiological values. ICP is only one of the main determinants of outcome in aSAH and therefore its independent contribution may need to be studied together with other signals or indexes that may detect other forms of brain-threatening events. In this context, it would be of great interest to assess the association between ICP and outcome in relation to the CAR status of the patient. In patients with aSAH, impaired autoregulation is associated with DCI and poor long-term outcomes [16, 17]. Similarly, it could be of interest to integrate the ICP with local measurements of brain oxygenation ($PbtO_2$). $PbtO_2$ can be used as a monitor of tissue hypoxia due to DCI [18], which is a main contributor to outcomes in aSAH. In fact, extremely reduced $PbtO_2$ can occur also in absence of elevated ICP, leading to poor short and long-term outcomes [19]. To perform these studies, large multi-center databases of high-quality monitoring data are needed. However, this may be one of the main challenges of aSAH research. The uncertainty that revolves around aSAH monitoring and treatment, resulted in a high variability in treatment protocols among centers. This variability makes large-scale research based on multi-center collaborations challenging, which in turn further fuels the lack of knowledge on aSAH pathophysiology.

As mentioned before, ICP monitoring for severe aSAH is not standard practice in most ICUs. Invasive ICP monitoring has an intrinsic risk for the patient, due to infections, brain tissue lesions or hemorrhage. In addition, it is a costly technique that requires specialized personnel and a neurosurgical setting. For some centers, the benefits of invasive ICP monitoring in patients with aSAH do not outweigh the related risks and

costs, especially considering the lack of specific recommendations on ICP management. Given these issues, non-invasive assessment of ICP could be a valuable solution for these patients. A precise enough technique to measure ICP non-invasively has not been developed yet. Several studies evaluated the possibility to measure ICP non-invasively with the transcranial doppler, optic nerve sheath diameter or imaging based sensors but with limited success [20].

9.3 DEVELOPMENT OF DECISION SUPPORT TOOLS FOR THE MANAGEMENT OF SEVERE TRAUMATIC BRAIN INJURY

9.3.1 Main findings

Severe TBI is a critical medical condition, with high mortality rate and high prevalence of long-term neurological impairment among survivors [21, 22]. In TBI, outcomes are mostly determined by the occurrence of secondary brain injuries (SBI) which are caused by intracranial and systemic events that follows the initial traumatic injury. To optimize outcomes, management of TBI mostly aims at keeping the patient stable while avoiding SBI. One of the leading mechanisms for SBI is the abnormal increase of the ICP of the patient, which can provoke mechanical distortion of brain tissues, midline shift or herniation, or could impair cerebral perfusion.

While the ICP plays a crucial role in TBI pathophysiology, guidelines for the optimal treatment strategy for elevated ICP are mainly expert-based [23]. Current Brain Trauma Foundation guidelines suggest a threshold-based approach that expects to start aggressive treatment when the ICP rises above 22 mmHg [24]. This approach evaluates ICP harmfulness by sorely considering the ICP absolute value, although several studies suggested that the combination of ICP intensity and changes of ICP over time, quantified with the ICP dose, may better evaluate the potential risk of an ICP event. The association between ICP dose and outcomes could be visualized with the method proposed by Güiza et al. [13] who demonstrated that events of ICP dose that occur more frequently in patients with better and worse long-term outcomes, could be represented in two separated areas. In addition, several factors may contribute to the ability of the patients to tolerate abnormal ICP, for example the CAR status of the patient should be taken into consideration when evaluating therapy in the presence of elevated ICP. In clinical practice, clinical decision making based on CAR is challenging,

because of the lack of validated methods to assess or quantify CAR at the bedside.

To assist decision-making in TBI management and promote a more proactive treatment of elevated ICP, Güiza et al. [25] proposed a prediction model for the early-detection of events of extremely elevated ICP in patients with TBI. The model predicted with a 30 minute forewarning events of ICP > 30 mmHg that lasted more than 10 minutes. One important aspect for the translation into clinical practice is the external validation of these models on multi-center, large, unseen data. External validation quantifies the ability of the model to generalize from the development data (data used to learn a task) and successfully perform a task when applied to unseen data. In Chapter 5 we externally validated the prediction model of Güiza et al. [25] on the High-Resolution CENTER-TBI dataset. When evaluated on this large, multicenter European dataset of prospectively collected data of patients with severe TBI, the model presented robust performance as compared to the performance on the original development cohort. The results presented in Chapter 5 demonstrate the generalizability capacities of the model, with performance that was unaffected by potential changes in clinical practice or clinical settings.

Beyond the ICP events that were targeted by the prediction model of Güiza et al. [25] , there are other ICP events, of lower intensity but longer duration, that are equally associated with poor long-term outcomes. Therefore, in Chapter 6 we further worked on the concept of ICP prediction, but differently from what was done in previous studies, we did not focus on the prediction of one specific ICP event, of specific intensity and duration, but on the prediction of a broad range of ICP doses that have been associated with poor long-term outcomes in previous studies [13]. The definition of the prediction target based on the "dose" represents the novelty of this prediction model. Specifically, the prediction target included a broad range of doses of elevated ICP that have been associated with poor neurological outcomes in the visualization of Güiza et al. [13]. The model was developed on a multi-center dataset of 290 adult patients with severe TBI and externally validated on 264 patients from the High-Resolution CENTER-TBI dataset. The prediction model presented in Chapter 6 presented good and robust performance on the development and external validation set, providing timely and accurate predictions of a broad range of harmful ICP doses. The model can be a valuable alerting tool for when a

patient is at high risk of harmful ICP doses and therefore requires medical evaluation.

To promote the use of this prediction model in future interventional studies, in Chapter 7 we developed a prototype software that displays the model predictions as well as other potentially relevant information for decision support and treatment optimization of patients with TBI. Although the software does not suggest therapeutic intervention, we can still talk about decision support given that it informs the clinician on when a clinical decision may be required. In detail, the software displays the continuous predictions of harmful ICP events, the LAx as index of CAR, and the real-time visualization of the ICP and LAx intensity-duration burden. The interface of the software was developed in close collaboration with the nurses and clinicians of the ICU of the University Hospitals of Leuven (UZ Leuven). In addition, the prototype software was integrated and tested in a blinded mode in the ICU of UZ Leuven. This study allowed us to obtain a robust and interpretable tool for future interventional studies. Moreover, demonstrating accuracy and clinical utility on retrospective data, does not guarantee that the model will perform equally well when applied to real-time clinical data. Therefore, in Chapter 8 we performed a preliminary evaluation of the performance of the prediction model for harmful ICP doses (presented in Chapter 6) when applied to continuous data prospectively collected at the bedside. When evaluated on these real clinical data, the model presented excellent discrimination and specificity, good calibration, and clinical usefulness for a broad range of risk thresholds. The performance of the model on this small dataset justifies a future interventional study. Future interventional studies are required to assess whether the use of this model at the bedside could improve patient's outcomes. For the successful execution of these interventional studies, this model needs to be integrated into clinical care workflow and be available as usable and interpretable tool, therefore the importance of the development of the prototype software that was presented in Chapter 7.

9.3.2 Current impact of research and future perspectives

New decision-support tools for the management of patients with severe TBI could be used to aid clinicians in the challenging task of interpreting neuromonitoring values to steer therapy. In this thesis, we developed such software for treatment optimization in TBI. The research presented in this thesis covers most of the necessary steps of software development, from

model development to external validation and proof-of-concept in a small clinical trial. The result is a prototype software that can be used in a future interventional study.

One of the main features of this software is the implementation of the prediction model for harmful doses of ICP, which was presented in Chapter 6. The use of this prediction model at the bedside may contribute to overcoming the limitations of the current standard management of elevated ICP. The model allows for early interventions against the entire spectrum of doses of ICP that have been associated with outcomes that define a poor neurological status. This is an innovative approach, not only because it reformulates the treatment target from prevention of ICP elevations to prevention of harmful events of ICP doses, but also because the formulation of the prediction target explicitly strives for better long-term outcomes.

The prediction model of Chapter 6 predicts a broad range of ICP doses in the "red area" of the visualization plots of Güiza et al. [13]. A future extension of this prediction model may focus on identifying the exact position of the patient in the red area of the visualization plot, in the next 30 minutes. This information would be complementary to the alerts provided by the model presented in Chapter 6 and it may help the clinician to better tailor the clinical intervention in response to an alert from the prediction model for harmful ICP doses. In addition, future versions of the prediction model presented in Chapter 6 should define the targeted red area in relation to constant changes in CAR status. In fact, it was demonstrated that the color-coded visualization from Güiza et al. [13] changes in relation to the CAR status of the patient (quantified with the LAx) [13], showing that in case of impaired CAR the ability of the patient to tolerate elevated ICP decreases. In a future version of the prediction model, these more complex versions of the visualization plots should be considered. In addition, it would be interesting to combine the predictions of the model for future doses of ICP with those of prediction models for different types of future brain-threatening events. For example, it would be interesting to predict events of reduced $PbtO_2$ or reduced cardiac output, and then combine them with the prediction model of Chapter 6 not only to get more accurate predictions of impeding harmful ICP doses, but also to detect a broader range of secondary brain injuries or early signs of reduced intracranial compliance. The model developed by Myers et al. [26] is a promising example of a prediction model for $PbtO_2$ crisis in TBI, it presented good accuracy, sensitivity and specificity on the external validation dataset (only

temporal validation). However, these results need to be validated on a multicenter external dataset. To the best of our knowledge accurate models for prediction of cardiac output or heart rate in TBI are not available.

Moreover, it is important to mention that the harmfulness of doses of ICP may depend on the remaining capacity of the brain to tolerate (or compensate) intracranial hypertension. In view of individualized ICP treatment, brain tolerance to intracranial hypertension should be taken into consideration when evaluating the treatment strategy of the patient when reacting on an alert from the model. Brain tolerance can be roughly estimated with different complementary methods. For example, by evaluating the CAR status of the patient, by monitoring local cerebral oxygenation with the PbtO2, by examining the pupillary reactivity or by assessing cerebral blood flow through Transcranial Doppler. Although relevant for the evaluation of the clinical status of the patients, all these methods have important limitations and a safe and effective protocol that explicitly includes the evaluation of brain tolerance in the management of elevated ICP is still missing.

The impact of the use of the prediction model for harmful doses of ICP in clinic practice, on patient's outcomes will need to be evaluated with an interventional study. In such a study, the model will provide alerts to the clinician of when a patient is at risk of harmful ICP doses and therefore in need of special medical attention, however, the definition of the clinical strategy of the patient will remain at total discretion of the clinician. To avoid bias towards non-monitored patients, the included patients will be all monitored with the decision-support tool. The time of the alerts will be recorded as well as the administered treatment, if any. Short term and long-term outcomes, such as length of ICU stay, mortality and Glasgow Outcome Scale at 6 months, will be collected and later compared with an historical control group, for example the CENTER-TBI. In addition, prevalence of elevated ICP and ICP in the red area will be calculated and compared between the intervention and the historical control group. KU Leuven internal funding has been granted (C3/21/071) to fund this project, which should start in the second trimester of 2022. Importantly, short- and long-term outcomes in TBI are not only determined by elevated doses of ICP, but also affected by the intracranial underlying mechanisms, such as excitotocitity, mitochondrial dysfunction or oxidative stress that also provoke a rise in ICP. Therefore, it remains unclear to which extent patients'

outcomes may improve by only treating elevated dose of ICP. The planned interventional study may partially answer this question.

In this thesis, we implemented a prototype decision-support tool that displays prediction of harmful ICP events (from the machine learning models presented in Chapter 5 and 6) so as other relevant information. The software has been tested at the bedside on real clinical data (Chapter 7 and 8). This type of study is unique, in fact despite the surge in machine learning (ML) models for clinical problems and the rising expectations on the use of AI and ML at the bedside, very few examples of ML models have actually made it to the implementation phase. One of the main barriers in the translation of ML models at the bedside are the several ethical and regulatory issues that result from the application of relatively new statistical methods to a highly regulated field. Moreover, it is crucial that these models address specific clinical needs, in a clinically relevant way, but also that they are trusted and well accepted by nurses and clinicians. Trust can be gained by transparent and effective reporting to facilitate understanding and replicability. Moreover, it is indispensable to involve nurses and/or clinicians during the entire process of model development and implementation at the bedside and allow them be the main drivers of ML research in healthcare. The promotion of interdisciplinary training for the new generation of clinicians could provide a better overview of the methodology behind ML models with several benefits. First, promote a closer collaboration between clinicians and data scientists. Second, allow the clinician to have a critical viewpoint towards ML tools and recognize value beyond the hype. Third, make the clinicians more aware of the strengths and limitations of this new methodology, promoting a better interpretation of the results of ML research on clinical data. Fourth, once a ML model is implemented at the bedside, increased awareness will allow for a constructive computer-doctor collaboration. The ML model will become an additional diagnostic/decision-support/alerting tool at disposition of the clinician, who will be able to interact with the model by setting the best alerting threshold depending on whether type I error (false positives) or type II error (false negatives) needs to be avoided. This interdisciplinary trend has already started, as an example of this, the recent editions of the European Society of Intensive Care Medicine (ESICM) conferences included one thematic program that was dedicated to AI and ML in the ICU. Another barrier against the implementation of ML at the bedside are the several technical challenges. In fact, beyond the technical

challenge of implementing the bedside software itself, testing a prototype software at the bedside requires the close collaboration with the IT department of the hospital and the compliance with all the safety regulations of the hospital on the acquisition and processing of medical data. Of particular interest would be the development of a common research platform based on the collection of data from the hospitals electronic health records where researchers can easily implement their AI model and test it in a blinded mode in real clinical setting. This type of collaboration would facilitate the multicenter, blinded, prospective evaluation of these models at the bedside for those research groups with little background in data science or programming.

In Chapter 8 of this thesis, we evaluated the performance of the model for the prediction of harmful ICP doses in real-time clinical data. The evaluation of the real-time performance of the model with standard performance metrics such as accuracy, sensitivity or precision is challenging. These metrics may give an idea of the performance of the model, but do not take into consideration the trend over time of the predictions and do not integrate considerations that could be assessed only visually. The definition of new performance evaluation metrics would be of great interest for the scientific community and would probably provide additional instruments for the evaluation of these models on real clinical data.

This research was made possible by using high-quality, large, prospectively collected, multi-center, open-source datasets of clinical and monitoring data of patients with TBI. These datasets are the only mean by which impactful research is possible and common scientific efforts dedicated to the collection of such datasets should continue. The collection of multicenter datasets is challenging and requires a strong collaboration between parties. One of the barriers in the collection of multi-center datasets is the aggregation of data from disparate sources, and the lack of a uniform storage protocol across centers. Moreover, the relatively new General Data Protection Regulation (GDPR) had a significant impact on healthcare organization and data collection. The GDPR contemplates for health data a higher level of protection and as such requires collecting explicit consent from each patient or their relative. In addition, given the relatively low incidence of severe TBI, collection of big datasets of patients with TBI is a long and resource-intensive task. The collection of high-quality dataset requires big investments, which may be even more

challenging given that funding for TBI research is limited when compared to other conditions like cancer.

9.4 GENERAL CONCLUSIONS

In this thesis, we used AI to gain insights on the role that neuro-monitoring signals can play in the optimization of treatment in critically ill patients at risk of brain injuries. First, we demonstrated that in children with CHD, post-operative $SctO_2$ desaturation independently increases the probability of lower total IQ 2 years after PICU admission. Desaturation was demonstrated to be an important clinical event in the post-operative management of children with CHD. Future interventional studies need to assess whether management strategies aimed at preventing or treating post-operative desaturation will result in better outcomes. Second, we investigated the association between elevated ICP doses and long-term neurological outcomes in patients with aSAH. We demonstrated that the association between ICP dose and long-term outcomes can be visualized in a color-coded plot and that the dose of ICP may have a greater clinical value than the ICP absolute value. Third, we demonstrated that a previously developed prediction model for elevated ICP has robust performance when evaluated on a large, multicenter dataset. Fourth, we developed and validated an accurate prediction model for harmful doses of ICP in patients with severe TBI. The model presented good and robust performance when evaluated on the development and external validation dataset, demonstrating clinical usefulness for a broad range of risk thresholds. Moreover, to prepare for a future clinical implementation, we developed a prototype software that implements such prediction model so as previous research results and we tested it at the bedside. Finally, we prospectively evaluated the performance of the model for the prediction of harmful ICP doses on a real clinical setting, where the model demonstrated good enough performance to proceed to an interventional study.

BIBLIOGRAPHY

1. Flechet M, Güiza F, Scharlaeken I, et al (2019) Near-Infrared–Based Cerebral Oximetry for Prediction of Severe Acute Kidney Injury in Critically Ill Children After Cardiac Surgery. Crit Care Explor 1:e0063. https://doi.org/10.1097/CCE.0000000000000063

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

3. Kussman BD, Wypij D, Laussen PC, et al (2010) Relationship of intraoperative cerebral oxygen saturation to neurodevelopmental outcome and brain magnetic resonance imaging at 1 year of age in infants undergoing biventricular repair. Circulation 122:245–254. https://doi.org/10.1161/CIRCULATIONAHA.109.902338

4. Hoffman GM, Brosig CL, Mussatto KA, et al (2013) Perioperative cerebral oxygen saturation in neonates with hypoplastic left heart syndrome and childhood neurodevelopmental outcome. J Thorac Cardiovasc Surg 146:1153–1164. https://doi.org/10.1016/j.jtcvs.2012.12.060

5. Deschamps A, Lambert J, Couture P, et al (2013) Reversal of Decreases in Cerebral Saturation in High-Risk Cardiac Surgery. J Cardiothorac Vasc Anesth 27:1260–1266. https://doi.org/10.1053/j.jvca.2013.01.019

6. Vesoulis ZA, Lust CE, Liao SM, et al (2016) Early hyperoxia burden detected by cerebral near-infrared spectroscopy is superior to pulse oximetry for prediction of severe retinopathy of prematurity. J Perinatol 36:966–971. https://doi.org/10.1038/jp.2016.131

7. Alderliesten T, Lemmers PMA, Smarius JJM, et al (2013) Cerebral Oxygenation, Extraction, and Autoregulation in Very Preterm Infants Who Develop Peri-Intraventricular Hemorrhage. J Pediatr 162:698-704.e2. https://doi.org/10.1016/j.jpeds.2012.09.038

8. Sola A, Golombek SG, Montes Bueno MT, et al (2014) Safe oxygen saturation targeting and monitoring in preterm infants: can we avoid hypoxia and hyperoxia? Acta Paediatr 103:1009–1018. https://doi.org/10.1111/apa.12692

9. Dix LML, van Bel F, Lemmers PMA (2018) Monitoring Cerebral Oxygenation in Neonates: An Update. Front Pediatr 5:1–9. https://doi.org/10.3389/fped.2017.00160

10. Zoerle T, Lombardo A, Colombo A, et al (2015) Intracranial pressure after subarachnoid hemorrhage. Crit Care Med 43:168–176. https://doi.org/10.1097/CCM.0000000000000670

11. Magni F, Pozzi M, Rota M, et al (2015) High-Resolution Intracranial Pressure Burden and Outcome in Subarachnoid Hemorrhage. Stroke 46:2464–2469. https://doi.org/10.1161/STROKEAHA.115.010219

12. Cossu G, Messerer M, Stocchetti N, et al (2016) Intracranial pressure and outcome in critically ill patients with aneurysmal subarachnoid hemorrhage: A systematic review. Minerva Anestesiol 82:684–696

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

14. Donnelly J, Güiza F, Depreitere B, et al (2021) Visualising the pressure-time burden of elevated intracranial pressure after severe traumatic brain injury: a retrospective confirmatory study. Br J Anaesth 126:e15–e17. https://doi.org/10.1016/j.bja.2020.09.018

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

16. Budohoski KP, Czosnyka M, Kirkpatrick PJ, et al (2013) Clinical relevance of cerebral autoregulation following subarachnoid haemorrhage. Nat Rev Neurol 9:152–163. https://doi.org/10.1038/nrneurol.2013.11

17. Budohoski KP, Czosnyka M, Smielewski P, et al (2012) Impairment of Cerebral Autoregulation Predicts Delayed Cerebral Ischemia After Subarachnoid Hemorrhage. Stroke 43:3230–3237. https://doi.org/10.1161/STROKEAHA.112.669788

18. Helbok R, Madineni RC, Schmidt MJ, et al (2011) Intracerebral Monitoring of Silent Infarcts After Subarachnoid Hemorrhage. Neurocrit Care 14:162–167. https://doi.org/10.1007/s12028-010-

9472-9

19. van den Brink WA, van Santbrink H, Steyerberg EW, et al (2000) Brain Oxygen Tension in Severe Head Injury. Neurosurgery 46:868–878. https://doi.org/10.1097/00006123-200004000-00018

20. Harary M, Dolmans RG, Gormley W (2018) Intracranial Pressure Monitoring—Review and Avenues for Development. Sensors 18:465. https://doi.org/10.3390/s18020465

21. Maas AIR, Menon DK, Adelson PD, et al (2017) Traumatic brain injury: integrated approaches to improve prevention, clinical care, and research. Lancet Neurol 4422:. https://doi.org/10.1016/S1474-4422(17)30371-X

22. Steyerberg EW, Wiegers E, Sewalt C, et al (2019) Case-mix, care pathways, and outcomes in patients with traumatic brain injury in CENTER-TBI: a European prospective, multicentre, longitudinal, cohort study. Lancet Neurol 18:923–934. https://doi.org/10.1016/S1474-4422(19)30232-7

23. Hawryluk GWJ, Aguilera S, Buki A, et al (2019) A management algorithm for patients with intracranial pressure monitoring: the Seattle International Severe Traumatic Brain Injury Consensus Conference (SIBICC). Intensive Care Med 45:1783–1794. https://doi.org/10.1007/s00134-019-05805-9

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

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

26. Myers RB, Lazaridis C, Jermaine CM, et al (2016) Predicting Intracranial Pressure and Brain Tissue Oxygen Crises in Patients with Severe Traumatic Brain Injury. Crit Care Med 44:1754–1761. https://doi.org/10.1097/CCM.0000000000001838

SUMMARY

Critically ill patients admitted in an intensive care unit (ICU) may suffer from long-term neurological impairment with important repercussions on their quality of life after ICU discharge. The prevention of neurological damage is a main concern in critical care and in some patients the clinical status of the brain is continuously monitored through several modalities. However, the large amount of generated data may be difficult to interpret and as a result, subtle patterns of occurring neurological damage may remain unnoticed. More advanced methods to process and interpret these data may be needed. In this context, artificial intelligence (AI) can be used to perform patter recognitions, classification or predictions. More generally, AI indicates a group of statistical and mathematical methods that allow to perform a task automatically and therefore enable problem-solving and decision making.

The general objective of this PhD thesis was to use advanced AI techniques to gain insights on the association between neuro-monitoring signals and long-term outcomes in patients at risk of brain injury and to develop new tools for treatment optimization in neuro-monitored patients. This research project consisted in five parts, where we focused on three different types of patients at risk of brain injury, namely children after cardiac surgery for congenital heart defects, patients with sub-arachnoid hemorrhage and patients with traumatic brain injury.

In the **first part** of this research project, we studied the independent association between reduced cerebral tissue oxygen saturation ($SctO_2$) as measured by the near-infrared spectroscopy (NIRS), and long-term neurocognitive outcomes in children that underwent surgery for congenital heart defects (CHD). Children with severe CHD can suffer from reduced cerebral oxygenation, which can cause hypoxic and/or ischemic damage to the brain and lead to long-term neurodevelopmental deficits. Therefore, during the entire perioperative period, local SctO2 of the frontal lobe of the child is monitored non-invasively with the NIRS. However, physiological values of $SctO_2$ are not well defined and consequently specific $SctO_2$-based guidelines for treatment optimization are missing. In this research project,

we gained insights on the role of $SctO_2$ as neuro-monitoring signal in a pediatric setting. We used prospectively collected data of children after cardiac surgery for CHD to investigate the association between $SctO_2$ and total IQ at two years follow-up. We found that increased $SctO_2$ desaturation (defined as $SctO_2$<65%) and decreased $SctO_2$ mean in the early-postoperative period independently increased the probability of a lower total IQ at 2-years follow-up. These results suggest that $SctO_2$ as measured by the NIRS could be a valuable neuro-monitoring signal for the post-operative management of children with CHD. Future studies should investigate whether a NIRS-based protocol that targets $SctO_2$ values between 65% and 85-90% would result in improved long-term neurocognitive outcomes.

In the **second part** of this research project, we studied the independent association between doses of elevated intracranial pressure (ICP) and neurological outcomes, in patients with aneurysmal sub-arachnoid hemorrhage (aSAH). Patients with aSAH may suffer from elevated ICP, with potential consequences on their long-term neurological outcomes. Although elevated ICP in patients with aSAH is associated with poor outcomes, specific recommendations for its treatment are missing and current clinical practice is mostly based on the guidelines for traumatic brain injury (TBI). In patients with TBI, the combination of intensity and duration of an ICP event, also called ICP dose, has been demonstrated to be associated with outcomes. More insights on the role of abnormal ICP in aSAH may provide the basis for treatment optimization, with a potential impact on long-term outcomes. In this project, we visualized the association between doses of elevated ICP and neurological outcome (6-month Glasgow Outcome Score [GOS]) in 2 groups of patients from 2 European ICUs. For both cohorts, the resulting color-coded plot showed an association between doses of ICP and 6-month GOS. A semi-exponential transition curve defined two areas of ICP doses that were associated with good or poor outcomes, indicating that the combination of duration and intensity of elevated ICP may play a role in the definition of outcomes in patients with aSAH. The cumulative time that the patient spent in the area of the visualization curves that was associated with poor outcomes resulted independently associated with poor 6-month GOS in a multivariable logistic regression model which was corrected for known confounders. The time spent in the red area was a stronger predictor of outcomes that the cumulative time with ICP above 20 mmHg. These results suggest that the

intensity and duration of an ICP events, i.e. the ICP dose, is a strong predictor of long-term neurological outcomes in patients with aSAH. Further multicenter studies are needed to validate these findings.

In the **third part** of this research project, we performed an external validation of a prediction model for extremely elevated ICP in patients with TBI. External validation is a crucial step to bring machine learning (ML) models at the bedside. It assesses the generalizability capacities of the model, and test whether the model has similar performance when applied on data from different settings or different periods of time. In this research project, we externally validated a ML model for the prediction of events of ICP above 30 mmHg. For the validation we used the High-Resolution Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (HR CENTER-TBI) dataset, a large multi-center dataset of prospectively collected data of patients with severe TBI that were admitted in the ICUs of 35 European centers between 2015 and 2017. On this external dataset, which was collected almost 10 years later than the original development cohort, the model presented robust performance, with excellent accuracy, sensitivity and specificity and good calibration. Moreover, the model demonstrated clinical usefulness for all clinical relevant risk-thresholds. The results demonstrate that this prediction model is robust to inter-center variability and potential changes in clinical practice.

In the **fourth part** of this research project, we developed and externally validated a model for the prediction of potentially harmful doses of ICP in patients with severe TBI. In patients with TBI, prolonged elevated ICP is associated with poor long-term neurological outcomes and mortality. Therefore, monitoring and treatment of abnormal ICP is a cornerstone in the management of severe TBI. Current guidelines suggest a threshold-based approach, where the initiation of aggressive treatment is triggered by the rise of ICP above 22 mmHg, although the combination of intensity and duration of an ICP event, also called “ICP dose” may better describe the neurological risk that is associated with an ICP event. Prediction of future doses of ICP that have been associated with reduced long-term outcomes may allow for early intervention and treatment optimization in severe TBI. In this study, we developed and validated a ML model that predicts with a 30 minutes forewarning events of ICP doses that have been associated with poor neurological outcomes in previous studies. For this scope, we used as development and external validation datasets two large European multi-

center datasets of monitoring data of patients with TBI. The prediction model provided accurate predictions of potentially harmful doses of ICP on the development and external validation dataset. Moreover, in both datasets the model demonstrated clinical usefulness for a broad range of risk thresholds. Whether early intervention on the basis of ICP dose predictions will result in improved short and long-term outcomes will be tested in a future interventional study. For this scope, it is important to first evaluate the performance of the model when used on a real clinical setting. This crucial passage evaluates the level of clinical readiness for bedside implementation of a model. Therefore, in the **fifth part** of this research project, we first developed a prototype software for the implementation of the prediction model for harmful ICP doses at the bedside. Later, we assessed the performance of the prediction model for harmful ICP doses on a real clinical setting. The prototype software was designed and developed in close collaboration with nurses and clinicians of the ICU of UZ Leuven and displays in an interpretable fashion the predictions of harmful ICP doses so as other research results that can provide valuable clinical information. The prototype was tested at the bedside in a blinded mode in a technical validation, which resulted in a robust software that can be used as research tool in an interventional clinical trial. Part of the data prospectively collected during the technical validation were used for the assessment of the model performance in real-time. As expected, the model presented a slight drop in performance when evaluated on continuous data, but still presented good calibration and clinical usefulness for a broad range of risk thresholds. Importantly, performance highly depended on the choice of the alerting threshold, which however will likely depend on the local policies of the center and on the personal needs and preferences of the patient. This study is a fundamental step for the future implementation of the model at the bedside since it demonstrates good enough performance to warrant to proceed to an interventional study and it already provides a functioning, robust software prototype for the scope.

In conclusion, the research presented in this thesis provided new knowledge on the role of neuro-monitoring signals in the early prediction of occurring neurological damage in patients with brain injuries. Moreover, we translated part of this new knowledge in a bedside prototype software, that opens avenues for the execution of an interventional study and the assessment of the impact that the use of these research results at the bedside may have on patients outcomes.

CURRICULUM VITAE

GIORGIA CARRA

BIOMEDICAL ENGINEER

Biomedical engineer with an expertise in the deployment of machine learning (ML) models in the intensive care unit.
I am very passionate about conceptualizing problems and translating research to bedside applications.

giorgia.carra@gmail.com
+32 485438658

EDUCATION

2017 – 2022 PhD student at KU Leuven, Belgium
FWO SB Fellowship
Title of PhD thesis: "Artificial intelligence for decision support and treatment optimization in patients with brain injuries"

2016 – 2017 Erasmus year at the University College of Cork, Ireland

2015 – 2017 Master in Biomedical Engineering, Polytechnic of Turin, Italy
Magna cum laude

2012 - 2015 Bachelor in Biomedical Engineering, Polytechnic of Turin, Italy

CORE COMPETENCES

- Artificial intelligence, statistic
- Research
- Project management
- Research valorization
- Scientific writing

TECHNICAL COMPETENCES

- Python (Sklearn, Keras, Kivy)
- SQL language (MySQL, MSSQL)
- R
- Matlab

SOFT SKILLS

- Presentation skills
- Interdisciplinary communication
- Interdisciplinary teamwork

LANGUAGES

MOTHER TONGUE	Italian
C1	English
B1	French

LIST OF PUBLICATIONS

2021 **Giorgia Carra**, Fabian Güiza, Ian Piper, Giuseppe Citerio, Andrew Maas, Bart Depreitere, Geert Meyfroidt on behalf of CENTER-TBI High-Resolution ICU (HR ICU) Sub-Study Participants and Investigators. *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*. Submitted for publication.

2021 **Giorgia Carra**, Fabian Güiza, Sofie Dietvorst, Bart Depreitere, Pieter Wouters, Liese Mebis, Geert Meyfroidt. *Real-time performance of a prediction model for harmful doses of intracranial pressure: a descriptive analysis*. Manuscript ready for submission.

2021 **Giorgia Carra**, Marine Flechet, An Jacobs, Sören Verstraete, Dirk Vlasselaers, Lars Desmet, Hanna Van Cleemput, Pieter Wouters, Ilse Vanhorebeek, Greet Van den Berghe, Fabian Güiza, Geert Meyfroidt. *Post-operative cerebral oxygen saturation in children after congenital cardiac surgery and long-term total IQ: a prospective observational study*. Critical Care Medicine, 2021 Jun 1; 49(6):967-976

2021 **Giorgia Carra**, Francesca Elli, Bogdan Ianosi, Marine Flechet*, Lukas Huber, Verena Rass, Bart Depreitere, Fabian Güiza, Geert Meyfroidt, Giuseppe Citerio, Raimund Helbok. *Association of dose of intracranial hypertension with outcome in subarachnoid hemorrhage*. Neurocritical Care. 2021 Jun. 34(3):722-730

2021 **Giorgia Carra**, Fabian Güiza, Bart Depreitere and Geert Meyfroidt on behalf of CENTER-TBI High-Resolution ICU (HR ICU) Sub-Study Participants and Investigators. *Prediction model for intracranial hypertension demonstrates robust performance during external validation on the CENTER-TBI dataset*. Intensive Care Medicine, 2021 Jan; 47(1):124-126

2020 **Giorgia Carra**, Jorge I.F. Salluh, Fernando José da Silva Ramos, Geert Meyfroidt. *Data-driven ICU management: Using Big Data and algorithms to improve outcomes*. Journal of Critical Care. 2020 Dec. 60:300-304

2019 Oksana Semenova, **Giorgia Carra**, Gordon Lightbody, Geraldine Boylan, Eugene Dempsey, Andriy Temko. *Prediction of short-term health outcomes in preterm neonates from heart-rate variability and blood pressure using boosted decision trees*. Computer Methods and Programs in Biomedicine. 2019 Oct. 180:104996

LIST OF ABSTRACTS FOR INTERNATIONAL CONFERENCES

2018 **EuroNeuro**, Brussels. *Association of duration and intensity of intracranial hypertension insults with outcome in subarachnoid hemorrhage: an observational study of two cities*. Poster presentation.

2019 **ISICEM**, Brussels. *Association of duration and intensity of intracranial hypertension insults with outcome in subarachnoid hemorrhage: an observational study of two cities*. Poster presentation.

2019 **ESICM**, Berlin. *Burden of intracranial hypertension in subarachnoid hemorrhage in relationship with the cerebrovascular autoregulatory status*. Oral presentation.

2019 **ICP conference**, Leuven. *Near-infrared spectroscopy in pediatric patients after cardiac surgery: association with neurocognitive outcomes assessed at 2-years follow-up*. Oral presentation.

2020 **ESICM**, online conference. *Post-operative near-infrared spectroscopy in pediatric patients after cardiac surgery: association with total IQ assessed at 2-years follow-up*. Poster presentation.

2020 **ESICM**, online conference. *External validation of a Gaussian Process model for the prediction of elevated intracranial pressure on the Center-TBI dataset*. Poster presentation.

INVITED ORAL PRESENTATIONS AND SCIENTIFIC OUTREACH

2021 **Critical Care Medicine Journal Club**. *Postoperative Cerebral Oxygen Saturation in Children After Congenital Cardiac Surgery and Long-Term Total Intelligence Quotient: A Prospective Observational Study*.

2021 **ESICM,** online conference. *Can we make neuromonitors smarter?*

DATATHONS

2019 Participation to the 2019 edition of the Milan Critical Care Datathon. Member of the winning team, title of the project: "*Prediction of blood lactate kinetics through a machine learning algorithm in a large clinical data sets*"

2019 Mentor to the ICP conference Datathon

COURSES AND CERTIFICATES

2018	ACDL 2018 : Advanced Course on Data Science & Machine Learning (8 ECTS)
2018	ASPP: Advanced Scientific Programming in Python (2 ECTS)
2018	Geilo Winter School in eScience (60 hours ~1 ECTS)
2019	Good Clinical Practice course
2021	Interdisciplinary Program in Healthcare Innovation (5 ECTS)
2021	Research to market programme (30 hours ~1 ECTS)

KU LEUVEN TRAININGS

2018	Central lecture on research integrity
2018	Presentation and seminar skills for biomedical researchers
2018	Exploitation of Research and Technology & Knowledge Transfer
2018	Writing biomedical research papers, Intensive Academic writing
2018	Managing your PhD
2019	Python as second language, Python software engineering,
2019	GPU deep learning, GPU for programmers, GPU general concepts software applications
2020	Discover your career profile, Insights in your strengths and ambitions
2020	Introduction to leadership

ACKNOWLEDGMENT

This thesis is not only the result of my work, but also the result of the collaboration with several brilliant and visionary people that I want to thank from the bottom of my heart. Thank you to every person with who I brainstormed, exchanged ideas, discussed over opposite opinions, shared words of support. It was all very precious to me.

The biggest thank goes to my supervisors for teaching me how to perform high-quality scientific research, for encouraging me to accepting new challenges and for sharing their immense expertise with me. They are not only exceptional researchers, but also wonderful people. Geert, thank you for being my eye on the clinical word. Your leadership, enthusiasm and clinical expertise were essential for the successful execution of the project. Thank you for reminding me the importance of what we do. Fabian, I am glad to have worked with such an inspirational person. Thank you for always encouraging me to look a step (or two) forward. Thank you for your support, your empathy and for asking honestly interested questions about my well-being. Bart, thank you for the extremely valuable suggestions and clinical insights, which were always given with a natural kindness and humbleness.

A special thank also goes to professor Van den Berghe for creating such a multidisciplinary, positive and lively scientific environment. It was an honor to be part of it and to work with such a passionate and talented researcher.

I am grateful to my jury members. Professor Vens and professor Theys, thank you for following the progresses of my work and providing me very relevant feedback. I also want to thank professor Menon and professor Taccone for reviewing my thesis and for contributing to the scientific discussion during the PhD defense. To all my jury members, I am honored that my work has been reviewed by such brilliant researchers. A thank also goes to professor Milisen, who coordinated with mastery the reviewing process of this thesis, and to professor Agostinis for chairing the public defense.

This research could not have been possible without the support of the Fonds Wetenschappelijk Onderzoek (FWO), thank you FWO for funding this research. On this note, thank you KU Leuven for believing in our vision and funding the project that will follow this thesis.

A symbolic thank you to the patients that agree to participate to observational or prospective trials. For their gesture of altruism in a moment of difficulty.

For the NIRS-related research, thank you to Marine, Dirk, Lars and Sören for setting up the study and providing me feedback throughout the project. An additional thank goes to An, Hanna and Pieter for performing the long-term follow-up evaluation and collecting the follow-up data. Ilse, thank you for providing extremely valuable feedback on the manuscript. I am also grateful to everyone who contributed to the collection of the PEPaNIC database, without which part of my research would not had been possible.

For the SAH-related research, thank you to professor Citerio, professor Helbok and all the research and clinical team of the San Gerardo Hospital (Monza, Italy) and the University Hospital of Innsbruck (Innsbruck, Austria) for collecting the data and proving valuable feedback. Sofie, thank you for helping me with the interpretation of the results.

For the TBI-related research, a special thank you to the Brain-IT and CENTER-TBI investigators and collaborators for setting up such wonderful collaborations and making the collection of multicenter, high-quality monitoring data from patients with traumatic brain injury possible. Without such high-quality datasets our research would not have been possible. Thank you to Marine, for helping me in the initial (but crucial) phases of the TB-AI project. Thank you to the IT department of UZ Leuven, in particular to Mario Opsomer, Bart Vercammen, Sergey Balyaev for their incredible help during the prospective study on the TB-AI software. A special thank to Liese, Pieter, Sofie, Yuan, Sylvia, Sandra, Heidi for helping me with the execution of the TB-AI prospective study, without them it would have been impossible to reach such an important achievement.

I am very grateful for have found an amazing team of fellows PhD students: Yuan, Grégoire, Sofie, Marine, Bavo, Chloë, Lisa, An, Nathalie, Ruben, Arno, Wouter, Ines, Sam, Veerle, Astrid, Caroline and Lauren. Thank you for the always-positive mood of the lab, for the reciprocal help, for the exchange of ideas, for the honest suggestions and party times. You are such

a bunch of smart, easy-going, enjoyable people, and I am so glad to have met you all. Also, a great thank to Sarah, Sarah, Lies, Ilse, Inge and Lies for the nice chat and support! An additional thank you to Sarah for helping me with the organizational aspects of the thesis and defense, your kindness is heartwarming.

I would like to thank also all my friends and family that may have not contributed to the scientific content of this thesis, but certainly have had an important role in preserving my happiness and mental well-being in these last 4 years.

Il primo grazie di cuore va' alle mitiche *Party girls*, che sono state una parte importantissima di questo periodo. Grazie a Gaia, Fede, Marti, Manu, Je e Agnese per avermi fatto sentire a casa ad ogni cena del fine settimana, ogni volta che ci siamo prese il sole sul cemento del Rega, che abbiamo fatto un pigiama party sul divano Carra/D'Agostini, ogni volta che abbiamo bussato sotto un tavolo aspettando l'ansia, o abbiamo messo in pausa la nostra amicizia durante una sfida a Catan. Se ripenso a tutte le situazioni assurde e gioiose che abbiamo passato in questi anni non posso fare a meno di sorridere ed emozionarmi. Vi porto nel cuore.

Grazie ai miei amici di sempre, perché continuano ad esserci nonostante la distanza.

To Francisco, muito obrigada. You are my rock and I will always be grateful for the immense, unconditional support that you have given me in these past years (let's not talk about the last months!!). Thank you for the love and happiness that you give me every day. To this adventure, and to many to come.

Un ultimo ringraziamento va alla mia famiglia, a cui devo tutto e senza la quale non sarei qui a scrivere queste parole. Grazie mamma, papà e Sere per l'amore e il sostegno che mi date quotidianamente. Grazie perché ogni volta che ritorno a casa trovo lo stesso calore di sempre. Vi voglio un bene immenso.

Giorgia