Novel Methods to Predict Increased Intracranial Pressure During Intensive Care and Long-Term Neurological Outcome After Traumatic Brain Injury: Development and Validation in a Multicenter Dataset

Fabian Güiza, PhD¹; Bart Depreitere, MD, PhD²; Ian Piper, PhD³; Greet Van den Berghe, MD, PhD¹; Geert Meyfroidt, MD, PhD¹

Objective: Intracranial pressure monitoring is standard of care after severe traumatic brain injury. Episodes of increased intracranial pressure are secondary injuries associated with poor outcome. We developed a model to predict increased intracranial pressure episodes 30 mins in advance, by using the dynamic characteristics of continuous intracranial pressure and mean arterial pressure monitoring. In addition, we hypothesized that performance of current models to predict long-term neurological outcome could be substantially improved by adding dynamic characteristics of continuous intracranial pressure and mean arterial pressure monitoring during the first 24 hrs in the ICU.


Setting and Patients: The Brain Monitoring with Information Technology dataset consisted of 264 traumatic brain injury patients admitted to 22 neuro-ICUs from 11 European countries.

Interventions: None.

Measurements: Predictive models were built with multivariate logistic regression and Gaussian processes, a machine learning technique. Predictive attributes were Corticosteroid Randomisation After Significant Head Injury-basic and International Mission for Prognosis and Clinical Trial design in TBI-core predictors, together with time-series summary statistics of minute-by-minute mean arterial pressure and intracranial pressure.

Main Results: Increased intracranial pressure episodes in ICU could be predicted 30 mins ahead with good calibration (Hosmer-Lemeshow p value 0.12, calibration slope 1.02, calibration-in-the-large −0.02) and discrimination (area under the receiver operating curve = 0.87) on an external validation dataset. Models for prediction of poor neurological outcome at six months (Glasgow Outcome Score 1–2) based only on static admission data had 0.72 area under the receiver operating curve; adding dynamic information of intracranial pressure and mean arterial pressure during the first 24 hrs increased performance to 0.90. Similarly, prediction of Glasgow Outcome Score 1–3 was improved from 0.88 to 0.87 when including dynamic information.

Conclusion: The dynamic information in continuous mean arterial pressure and intracranial pressure monitoring allows to accurately predict increased CP episodes during intensive care. Adding information of the first 24 hrs of intracranial pressure and mean arterial pressure to known baseline risk factors allows very accurate prediction of long-term neurological outcome at 6 months. (Crit Care Med 2013;41:0–0)

Key Words: automated; data mining; decision support techniques; forecasting; Glasgow Outcome Scale; intensive care; intracranial hypertension; models; pattern recognition; statistical; traumatic brain injury

1 Department of Intensive Care Medicine, KU Leuven, Leuven, Belgium.
2 Department of Neurosurgery, KU Leuven, Leuven, Belgium.
3 Department of Clinical Physics, Southern General Hospital, Glasgow, Scotland, United Kingdom.

Supported by The European Framework-Programme (FP5-QLRI-2000-00454, QLGT-2002-00160 and FP7-IST-2007-217049); by the Foundation-for-Scientific-Research (in part) Flanders, Belgium (G.0904.11); and the Flemish government Methusalem-program (to G.V.d.B.).

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal’s Web site (http://journals.lww.com/ccmjournal).

The authors have not disclosed any potential conflicts of interest.
Drs. Meyfroidt and Güiza conceived and designed the study. Dr. Piper programmed the Brain-IT database, exported and validated the data. Drs. Piper and Depreitere are members of the Brain-IT steering group. Dr. Güiza programmed the machine learning algorithms and performed the statistic analysis. Drs. Meyfroidt and Güiza drafted the manuscript. Drs. Berghe, Depreitere, and Piper participated in the design of the study and proofread the manuscript. All authors read and approved the final manuscript.

For information regarding this article, E-mail: fabian.guiza@med.kuleuven.be

Copyright © 2013 by the Society of Critical Care Medicine and Lippincott Williams & Wilkins

DOI:10.1097/CCM.0b013e3182742d0a
Traumatic brain injury (TBI) is an important problem worldwide. In Europe, it is estimated that about 235 in every 100,000 people are hospitalized for or die from TBI every year (1). Over the past 30 yr, mortality from TBI has gradually declined to below 27% (2). This decline has been attributed to modern management of severe TBI, which is mainly concerned with prevention of secondary injury.

Monitoring of intracranial pressure (ICP) is recommended in every salvageable patient with severe TBI (defined as abnormal CT and a Glasgow Coma Scale [GCS] score of 3–8 (3)).

Patient studies have demonstrated a relation between increased ICP and poor outcome (4). However, severe ICP increases are often difficult to foresee. Early warning systems for this acute secondary injury would be useful in the management of severe TBI. They would expand the timeframe in which therapeutic measures could be taken, and therefore have the potential to prevent irreversible ischemic brain tissue damage. Additional intracranial monitoring systems have been proposed as a supplement to standard ICP monitoring. Some of these seem to precede or correlate with intracranial hypertension episodes, such as an increased lactate/pyruvate ratio obtained by microdialysis technology (5) or a worsening in brain tissue oxygenation (6). These monitoring systems have the disadvantage of being additionally invasive, labor-intensive and expensive. On the contrary, it is well-known that when forecasting a time-series, such as the ICP signal, past values are predictive of future values (7). A system that forecasts episodes of intracranial hypertension using only the standard monitoring after severe TBI, namely, ICP and mean arterial blood pressure (MAP), would be immediately applicable in many ICUs.

Some studies have identified dynamic characteristics of the ICP signal that seem predictive for outcome, such as the duration of increased ICP episodes (8), the area under the ICP curve (9, 10), the ICP variability (11, 12), and the CPP/ICP ratio (13). Despite this evidence for the predictive power of ICP dynamics, the available prognostic models remain based purely on static information (14). Currently, the International Mission for Prognosis and Clinical Trial design in TBI (IMPACT) (15) and Corticosteroid Randomisation After Significant Head Injury (CRASH) (16) predictive models are considered the golden standard. These multivariate logistic regression (LR) models are based on baseline static predictors and were developed and validated in very large databases.

Two important barriers hamper the use of dynamic physiological signals for predictive modeling. First, it requires collection and storage of high-quality validated data at high resolution for large populations. Second, methodologies to extract meaningful information from large amounts of high-dimensional data such as physiological time-series were hitherto unexplored. To address the first of these difficulties, the EU-funded Brain Monitoring with Information Technology (Brain-IT) (17) project was devised, resulting in a uniquely detailed database of TBI patients from neuro-ICUs across Europe. The second difficulty can be dealt with automatic knowledge extraction methods such as machine learning (18), which build prognostic models by exploring the entire database to identify relevant predictors and their interactions without prior clinical knowledge.

We hypothesize that the use of routinely monitored physiological signals allows for early, accurate prediction of acute secondary brain injury and long-term functional outcome in severe TBI patients. Our main interests are predictions that can impact clinical decision making for prevention of acute secondary brain injury and that can aid in counseling of long-term functional outcome after TBI. To this end, we present a model that predicts increased ICP episodes 30 mins in advance and models for early prediction of neurological outcome at 6 months. These novel models, based on machine learning techniques, incorporate dynamic information derived from ICP and MAP signals and are validated against state-of-the-art prognostic models.

**MATERIALS AND METHODS**

**Database**

The Brain-IT (17) database contains validated information of 264 TBI patients admitted to 22 neuro-ICUs in 11 European countries between March 2003 and July 2005. ICP-monitored patients were consecutively added to the database at each center. Among others, baseline risk factors, minute-by-minute ICP and MAP monitoring data, and Glasgow Outcome Score (GOS) at 6 months after the trauma are registered.

The use of these data for scientific purposes was approved by the Multi-Centre Research Ethics Committee for Scotland MREC/02/0/9 on February 14, 2002. All Brain-IT centers obtained local ethics or research committee approval and the need for informed consent was waived.

**Predictive Tasks**

Prediction of acute secondary brain injury comprised prediction of an increased ICP episode 30 mins in advance, at any time 4 hrs after admission. An episode is defined as an ICP above 30 mm Hg lasting at least ten consecutive minutes (Fig. 1). These values were chosen based on clinical judgment in the absence of a standard for duration in the definition of secondary insult. They are stricter than the EUSIG GRADE 2 definition for increased ICP (4) and are meant to more likely exclude transient phenomena such as increased ICP caused by coughing fits. A prediction horizon of 30 mins was deemed sufficient to allow for appropriate therapeutic intervention.

Early prediction of long-term functional outcome comprised the prediction, on admission or after 24 hrs, of poor neurological outcome, defined as GOS 1 (death) or GOS 2 (vegetative state) at 6 months. We chose to exclude severe disability (GOS 3) for our primary aim, under the assumption that only accurate prediction of GOS 1–2 would have clear consequences for clinical decision making. However, to allow comparison of our methodology with the existing CRASH and IMPACT models, we included the separate task of predicting unfavorable neurological outcome, defined as GOS 1, 2, or 3 at 6 months.
Study Cohorts

For prediction of increased ICP episodes, data from 239 patients who had ICP and MAP records lasting at least four consecutive hours anytime during ICU stay were used. Missing data were due to monitor disconnections or data loss during patient transport.

Complete data records were initially only available for 178 patients and later extended with data for the remaining 61 patients. The first set of patients was used as development cohort and the remaining as validation cohort. In total, 982 four-hour intervals with increased ICP episodes within the prediction horizon of 30 mins were identified in 108 (61%) patients in the development cohort. Likewise, there were 392 data instances with increased ICP episodes in 33 (54%) patients in the validation cohort. Nonevent data instances were also used; they were randomly selected 4-hr intervals for which no increased ICP episode occurred within the prediction horizon. This resulted in 2,677 total data instances for model development and 1135 for validation (Fig. 2A).

For early prediction of 6 months neurological outcome, patients for whom monitoring only began after the first 24 hrs or with unknown GOS because they were lost to follow-up were excluded. Data from the remaining 160 patients who had ICP and MAP records for the first 24 hrs of ICU stay and 6 months GOS were used. Of these patients, 29 (18%) had poor neurological outcome (GOS 1–2) and 69 (43%) had unfavorable outcome (GOS 1–3) at 6 months (Fig. 2B). Table 1 shows the demographic information for the different cohorts.

Predictors

Several techniques were used to analyze the ICP, MAP (and derived CPP) signals, and to generate the summary statistics referred to here as **dynamic predictors**. These included a) partitioning the time-series into nonoverlapping intervals and computing...
Güiza et al

Consort diagrams. GOS = Glasgow Outcome Score; ICP = intracranial pressure; MAP = mean arterial blood pressure.

Figure 2.

for each the median and a measure of signal variability; b) clustering the signals according to their overall shape for different intervals and resolutions; c) frequency-domain analyses; and d) correlations between ICP and MAP at different intervals. Examples of the techniques are shown in Figures 3 and 4, and a full list of the dynamic predictors for each task is available in Section A of the Electronic Supplement Material (Supplemental Digital Content 1, http://links.lww.com/CCM/A574). These analyses resulted in over a thousand potential dynamic predictors from which a subset of highly predictive and uncorrelated predictors were selected for each task via the Weka System (19) implementation of the correlation-based feature subset selection algorithm of (20).

For prediction of increased ICP episodes, the dynamic predictors were derived from the analysis of the selected 4-hr intervals of ICP and MAP. The ICU length of stay (LOS) at the time of prediction was also used as a predictor.

For early prediction of poor neurological outcome, two LR models based on static predictors (LR$_{GOS12-I}$ and LR$_{GOS12-C}$) and two GP models based on a combination of static and dynamic predictors (GP$_{GOS12-I}$ and GP$_{GOS12-C}$) were developed. LR$_{GOS12-I}$ and GP$_{GOS12-I}$ used the IMPACT-core predictors, while LR$_{GOS12-C}$ and GP$_{GOS12-C}$ used the CRASH-basic predictors. Similarly, four models were developed for early prediction of unfavorable neurological outcome (LR$_{GOS123-I}$, GP$_{GOS123-I}$, LR$_{GOS123-C}$, and GP$_{GOS123-C}$).

Validation

Prediction of increased ICP episodes was internally validated in the development cohort with ten repetitions of five-fold cross-validation. This validation was set up so that data instances from the same patient could not be used both for training and testing. Finally, a model built on the entire development cohort data was externally validated.

Early prediction of long-term neurological outcome was internally validated via 0.632+ bootstrapping (23), which is the preferred bootstrap technique for small datasets as it gives accurate and unbiased estimates of model performance (24). External validation was impossible because of insufficient data.

Predictor Ranking

The simple GPs with only two model parameters, used for prediction in this study, are essentially black-box models. More
### TABLE 1. Patient Demographic Information

<table>
<thead>
<tr>
<th></th>
<th>Episodes of Increased ICP</th>
<th>Long-Term Neurological Outcome</th>
<th>All Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Development Cohort (n = 178)</td>
<td>Validation Cohort (n = 61)</td>
<td>(n = 160)</td>
</tr>
<tr>
<td>LOS days, median (IQR)</td>
<td>14 (7–23)</td>
<td>16 (9.2–18)</td>
<td>14 (7.7–23)</td>
</tr>
<tr>
<td>Age, median (IQR)</td>
<td>33.1 (19–49)</td>
<td>24 (13–44)</td>
<td>33 (21–51)</td>
</tr>
<tr>
<td>Sex (% male)</td>
<td>80.9</td>
<td>77.1</td>
<td>79.4</td>
</tr>
<tr>
<td>Type of trauma (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Traffic accident</td>
<td>42.7</td>
<td>44.3</td>
<td>40.0</td>
</tr>
<tr>
<td>Fall</td>
<td>27.5</td>
<td>16.4</td>
<td>30.6</td>
</tr>
<tr>
<td>Pedestrian</td>
<td>7.9</td>
<td>9.8</td>
<td>7.5</td>
</tr>
<tr>
<td>Assault</td>
<td>9.0</td>
<td>11.5</td>
<td>8.1</td>
</tr>
<tr>
<td>Sport</td>
<td>3.4</td>
<td>4.9</td>
<td>2.5</td>
</tr>
<tr>
<td>Work</td>
<td>2.3</td>
<td>3.3</td>
<td>2.5</td>
</tr>
<tr>
<td>Unknown</td>
<td>5.6</td>
<td>9.8</td>
<td>6.2</td>
</tr>
<tr>
<td>GCS eye, median (IQR)</td>
<td>1 (1–3)</td>
<td>1 (1–2)</td>
<td>1 (1–2)</td>
</tr>
<tr>
<td>Missing (%)</td>
<td>2.8</td>
<td>6.5</td>
<td>3.1</td>
</tr>
<tr>
<td>GCS motor, median (IQR)</td>
<td>4 (2–5)</td>
<td>4 (2–5)</td>
<td>4 (2–5)</td>
</tr>
<tr>
<td>Missing (%)</td>
<td>1.7</td>
<td>6.5</td>
<td>1.9</td>
</tr>
<tr>
<td>GCS verbal, median (IQR)</td>
<td>1 (1–3)</td>
<td>1 (1–2)</td>
<td>1 (1–2)</td>
</tr>
<tr>
<td>Missing (%)</td>
<td>4.5</td>
<td>14.7</td>
<td>6.3</td>
</tr>
<tr>
<td>GCS total, median (IQR)</td>
<td>7 (4–10)</td>
<td>7 (4–9)</td>
<td>7 (5–10)</td>
</tr>
<tr>
<td>Missing (%)</td>
<td>5.6</td>
<td>14.7</td>
<td>7.6</td>
</tr>
<tr>
<td>Increased ICP episodes per patient, median (IQR)</td>
<td>1 (0–6)</td>
<td>1 (0–6)</td>
<td>1 (0–6)</td>
</tr>
</tbody>
</table>

ICP = intracranial pressure; LOS = length of stay; IQR = interquartile range; GCS = Glasgow Coma Scale.

complex models were built exclusively to generate a predictor ranking and thereby investigate the relevance of the individual predictors. These more complex models were GPs with one model parameter per predictor. A detailed description of the ranking process is available in Section C of the Electronic Supplement Material (Supplemental Digital Content 1, http://links.lww.com/CCM/A574).

**Evaluation Criteria**

Model discrimination was evaluated with the area under the receiver operating curve (AUROC) which is equivalent to the c-statistic. Calibration was evaluated with the Hosmer-Leme-show (HL) statistic, calibration-in-the-large, and calibration slope. Overall model performance was evaluated with the Brier score (BS) and Brier score scaled (BSS). A model was considered to discriminate well if its AUROC was above 0.8, and to be well calibrated if its HL p value was above 0.05, calibration-in-the-large close to 0, and calibration slope close to 1. Accurate models have a BS below the Brier\textsubscript{max} value for each task and a high BSS. Because the models are probabilistic, their accuracy, sensitivity, and specificity can be evaluated for different cutoff probabilities.

To determine whether differences in performance were statistically significant, the differences between models and corresponding bootstrap CIs were computed for each evaluation criterion. If the 95% CI did not include 0, then a statistically significant difference at the 0.05 level was declared. All analyses were done with the Weka System (19), the GP library (25), and the Statistics, Symbolic Math and Signal Processing toolboxes for MATLAB (The Math-Works, Natick, MA).

**RESULTS**

**Prediction of Increased ICP Episodes**

Table 2 reports model performance. In the development cohort, the model had good discrimination, calibration, and...
overall performance with 0.847 AUROC, 0.175 HL \( p \) value, 0.002 calibration-in-the-large, 0.99 calibration slope, 0.153 BS (below the 0.232 Brier\(_{\text{max}}\)), and 34.1% BSS.

Model performance remained unchanged in the validation cohort (Fig. 5), with 0.87 AUROC, 0.12 HL \( p \) value, −0.019 calibration-in-the-large, 0.14 BS, and 39.4% BSS; 77% classification accuracy, 82% sensitivity, and 75% specificity were obtained by applying the same probability cutoff that maximized sensitivity and specificity in the development cohort.

Besides ICU LOS at prediction time, 73 dynamic predictors were selected: 51 pertain to ICP (including absolute values, signal variability and frequency-domain coefficients), 14 to CPP (including absolute values for the first hour and frequency-domain coefficients), 5 to ICP-MAP correlations, and three to the most recent MAP variability. Exploration of predictor ranking demonstrated that most of the predictive power lies in the ICP signal proper, with the most recent measurements being more relevant.

**Early Prediction of 6 Months Neurological Outcome**

Table 3 reports performance for poor neurological outcome at 6 months. The \( \text{LR}_{\text{GOS}12} \) models performed poorly, with 0.72 AUROCs, calibration slopes below 0.81 indicating too extreme and optimistic predictions, BS equivalent to the 0.147 Brier\(_{\text{max}}\), and BSS below 7.7%. Accuracies, sensitivities, and specificities were between 69% and 73%. HL \( p \) values were above 0.43 and calibration-in-the-large close to 0.

The \( \text{GP}_{\text{GOS}12} \) models had excellent discrimination, calibration, and overall performance with AUROCs above 0.89, HL \( p \) values above 0.95, 0.93 calibration slopes, calibration-in-the-large close to 0, 0.08 BS (well below Brier\(_{\text{max}}\)), and BSS.
above 44%. Accuracies, sensitivities, and specificities were above 86%.

The models including dynamic information outperformed those based only on static predictors with a statistically significant difference (< 0.05) on all evaluation criteria except for calibration-in-the-large and calibration slope. There was no statistically significant difference in any performance criterion between \( \text{LR}_{\text{GOS12-I}} \) and \( \text{LR}_{\text{GOS12-C}} \) or between \( \text{GP}_{\text{GOS12-I}} \) and \( \text{GP}_{\text{GOS12-C}} \), indicating equal performance of the CRASH-basic and IMPACT-core predictors.

The \( \text{GP}_{\text{GOS12}} \) models used 19 dynamic predictors, of which six pertain to ICP, six to MAP, four to ICP-MAP correlations, two to CPP, and one to the number of increased ICP episodes. The highest ranked predictors were membership to specific curve clusters. Examples of highly ranked predictors are membership to the ICP curve cluster with normal admission values that consistently increased above 30 mm Hg by the end of the first 24 hrs (Fig. 4A) and to the MAP curve cluster indicating a pressure drop during the 22- to 24-hr interval (Fig. 4B). Signal variability was more predictive than absolute values. In addition, highly ranked was the evolution of the ICP-MAP correlation throughout the entire 24 hrs. The number of increased ICP episodes was ranked above the Glasgow motor score and above GCS and presence or extracranial injury in \( \text{GP}_{\text{GOS12-I}} \) and \( \text{GP}_{\text{GOS12-C}} \), respectively. Age was the highest ranked static predictor, positioned fourth in \( \text{GP}_{\text{GOS12-I}} \) and fifth in \( \text{GP}_{\text{GOS12-C}} \).

Table 4 reports performance for unfavorable neurological outcome at 6 months, which degrades slightly when compared with results for poor neurological outcome. The general trend, however, remains, with the models using static + dynamic predictors outperforming the models based only on static predictors, with statistically significant differences (<0.05) for all evaluation criteria except for the three calibration criteria. Performance of
the actual IMPACT-core and CRASH-basic models (via the online prognostic calculators) was very similar to that of LR
GOS123-I and LR
GOS123-C, demonstrating the accuracy of the validation technique used.

| TABLE 2. Performance of the Gaussian Process Model for Prediction of Increased Intracranial Pressure Episodes |
|-------------------------------------------------------|-------------------------------------------------------|-------------------------------------------------------|-------------------------------------------------------|
| Development Cohort (Data Instances = 2677) | Validation Cohort (Data Instances = 1135) |
| AUROC (c-statistic) | 0.847 | 0.872 |
| Hosmer-Lemeshow p value | 0.175 | 0.120 |
| Calibration-in-the-large | 0.002 | −0.019 |
| Calibration slope | 0.99 | 1.02 |
| Brier score | 0.153 | 0.137 |
| Brier score scaled (%) | 34.0 | 39.4 |
| Accuracy (%) | 77.1 | 77.4 |
| Sensitivity (%) | 79.2 | 81.6 |
| Specificity (%) | 76.1 | 75.2 |

AUROC = area under the receiver operating curve.
The last three rows were computed by choosing the cutoff that maximized sensitivity and specificity in the development cohort and then applying the same cutoff to the validation cohort.

\[ Brier\_max = incidence \times (1 - incidence)^2 + (1 - incidence) \times incidence^2. \]

\[ Brier\_max\ (development) = 0.232; \ Brier\_max\ (validation) = 0.226. \]

\[ Brier\ score\ scaled = 1 - (Brier\ score/Brier\_max). \]

The GP
GOS123-I models used 11 dynamic predictors, of which four pertain to ICP-MAP correlations, three to ICP, two to CPP, one to MAP, and one to the number of increased ICP episodes. The highest ranked predictors were ICP-MAP correlations and membership to specific curve clusters. The number of increased ICP episodes was again ranked above GCS and extracranial injury in the GP
GOS123-C model.

**DISCUSSION**

In this study, we presented models for prediction of acute secondary insults and long-term neurological outcome in severe TBI patients. First, we demonstrated that increased ICP episodes can be accurately predicted 30 mins in advance, based only on information derived from 4 hrs of ICP and MAP. Second, we showed that inclusion of dynamic information of ICP and MAP monitoring during the first 24 hrs in ICU outperforms the use of only static predictors for poor neurological outcome at 6 months.

| TABLE 3. Performance of Logistic Regression and Gaussian Process Models for Prediction of Poor Neurological Outcome (GOS 1 or 2) at 6 Mos |
|-------------------------------------------------------|-------------------------------------------------------|-------------------------------------------------------|-------------------------------------------------------|
| Poor neurological outcome (GOS 1, 2) at 6 months (n = 160) | | | |
| Static Predictors | Static + Dynamic Predictors | Static Predictors | Static + Dynamic Predictors |
| LR
GOS123-I (IMPACT-Core Predictors) | LR
GOS123-C (CRASH-Basic Predictors) | GP
GOS123-I (IMPACT-Core + Dynamic Predictors) | GP
GOS123-C (CRASH-Basic + Dynamic Predictors) |
| AUROC (c-statistic) | 0.72 (0.69–0.75) | 0.72 (0.68–0.74) | 0.90 (0.87–0.93) | 0.89 (0.86–0.92) |
| Hosmer-Lemeshow p value | 0.51 (0.27–0.71) | 0.43 (0.23–0.65) | 0.95 (0.92–0.97) | 0.95 (0.92–0.97) |
| Calibration-in-the-large | −0.003 (−0.030–0.031) | −0.003 (−0.030–0.026) | −0.008 (−0.027 to 0.015) | −0.004 (−0.026 to 0.020) |
| Calibration slope | 0.81 (0.60–1.05) | 0.72 (0.50–0.93) | 0.93 (0.82–1.04) | 0.94 (0.83–1.04) |
| Brier score | 0.14 (0.13–0.15) | 0.14 (0.13–0.15) | 0.08 (0.06–0.09) | 0.08 (0.07–0.10) |
| Brier score scaled (%) | 7.7 (2.5–11) | 5.6 (0–9.7) | 46 (37–53) | 44 (36–51) |
| Accuracy (%) | 70 (67–72) | 70 (66–72) | 88 (85–91) | 88 (84–91) |
| Sensitivity (%) | 73 (67–77) | 71 (66–76) | 88 (83–93) | 86 (83–91) |
| Specificity (%) | 69 (65–72) | 69 (64–73) | 88 (84–92) | 88 (83–92) |

GOS = Glasgow Outcome Score; LR = logistic regression; IMPACT = International Mission for Prognosis and Clinical Trial design in TBI; CRASH = Corticosteroid Randomisation After Significant Head Injury; AUROC = area under the receiver operating curve.

Median (25–75 percentiles). The last three rows were computed by choosing the cutoff that maximized sensitivity and specificity.

\[ Brier\_max = incidence \times (1 - incidence)^2 + (1 - incidence) \times incidence^2. \]

\[ Brier\ score\ scaled = 1 - (Brier\ score/Brier\_max). \]
experiments (data not shown) using simple dynamic predictors (ICP-MAP correlation, and frequency analysis). Additional power of more complex summary statistics (such as curve clustering) is an intuitive way of quantifying complex signal evolution and is extremely predictive. Third, most of the predictive power for increased ICP lies in the ICP signal itself. Fourth, the repeated selection of the ICP-MAP correlation as a relevant predictor hints at the importance of cerebrovascular autoregulation in the acute and chronic evolution of the TBI patient. Finally, in accordance with previous studies (12, 13), the variability of the signals is relevant for long-term predictions.

**CONCLUSION**

There is a wealth of untapped predictive power in the dynamical information of a patient’s routinely monitored physiological signals. Our data suggest that this information source is fundamental for accurate predictions of long-term neurological outcome and of acute secondary brain insults reflected by increased ICP. Future intervention studies are required to assess the impact of these predictions on patient outcome when used in clinical practice.

**ACKNOWLEDGMENTS**

We wish to acknowledge the Brain-IT steering members: Iain Chambers, Giuseppe Citerio, Bart Depreitere, Per Enblad, Barbara Gregson, Tim Howells, Karl Kiening, Julia Mattern, Pelle Nilsson, Ian Piper, Arminas Ragauskas, and Juan Sahuquillo.

---

*Figure 5.* Calibration plot for the increased intracranial pressure episodes prediction model in the validation cohort (n = 1135) with calibration slope 1.02 and calibration-in-the-large −0.019. Triangles indicate observed frequencies by deciles of predicted probability.

(GOS 1–2) or unfavorable (GOS 1–3) neurological outcome at 6 months.

To the best of our knowledge, this is the first prediction model proposed for increased ICP based purely on routinely collected ICP and MAP monitoring data. This presents an advantage over extrainvasive techniques investigated for this purpose such as microdialysis. Specifically, in a relatively small study (n = 25), hourly analysis of the lactate/pyruvate ratio predicted an ICP increase above 20 mm Hg within 3 hrs, with either 84% to 46% or 56% to 90% specificity-sensitivity (5). Although our results are not directly comparable as we are interested in a higher ICP threshold of 30 mm Hg, our model does have better discriminatory performance over the entire ROC space. In addition, because it only uses monitoring data, the increased ICP episode prediction model could be incorporated to bedside patient monitors currently in use as an early warning system. However, because the choice of probability cutoff trades off sensitivity and specificity, further analysis is required to evaluate the potential costs and risks of reacting to false alarms vs. the risk of leaving episodes undetected.

Prognostic models for long-term outcome must be extremely accurate if they are to be of real use to physicians in early clinical decision making. Consistent with early (26) and recent (8) studies, the key to obtaining this high accuracy lies in the use of information that captures the early clinical evolution of the patient. As our findings confirm, this can lead to outperforming models that rely exclusively on data elicited on admission (IMPACT and CRASH). Notably, this study demonstrates the high predictive power of more complex summary statistics (such as curve clustering, ICP-MAP correlation, and frequency analysis). Additional experiments (data not shown) using simple dynamic predictors (absolute values, averaging, and signal variation) in combination with static predictors did not result in statistically significant performance improvement over static predictors alone. Likewise, the performance of GP models based purely on static data (not shown) did not differ statistically from that of the LR models, indicating that the performance boost is due mainly to the addition of dynamic information.

Further investigation of monitoring data as a prognostic information source poses a challenge for future data collection in TBI, which we believe should invest in acquiring high-resolution time-series data for large patient populations. Exploration of these large databases is well suited for machine learning techniques, although an additional investment in high-end infrastructure to store the data and run the algorithms as well as specialized staff are necessary.

The major limitation of this study is population size. Although Brain-IT is one of the largest databases to include minute-by-minute monitoring data, the number of patients is small when compared with the databases used to develop and validate the IMPACT and CRASH models. As a consequence the long-term prediction models could not be externally validated. The use of robust internal validation techniques, however, effectively avoided overoptimistic performance evaluation.

A second shortcoming is the lack of an explanation for the underlying physiological mechanisms of the relevant dynamic predictors identified. Nevertheless, some observations can be made: first, the number of increased ICP episodes during the first 24 hrs in ICU is predictive of outcome at 6 months, suggesting of a link between early acute secondary brain insults and long-term neurological outcome. Second, curve clustering is an intuitive way of quantifying complex signal evolution and is extremely predictive. Third, most of the predictive power for increased ICP lies in the ICP signal itself. Fourth, the repeated selection of the ICP-MAP correlation as a relevant predictor hints at the importance of cerebrovascular autoregulation in the acute and chronic evolution of the TBI patient.
TABLE 4. Performance of Online Prognostic Calculators, Logistic Regression, and Gaussian Process Models for Prediction of Unfavorable Neurological Outcome (GOS 1, 2, or 3) at 6 Mos

<table>
<thead>
<tr>
<th>Unfavorable Neurological Outcome (GOS 1–3) at 6 Mos (n = 160)</th>
<th>Static Predictors</th>
<th>Static + Dynamic Predictors</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMPACT-Core Online Prognostic Calculator*</td>
<td>CRASH-Basic Online Prognostic Calculator*</td>
<td>LR_{GOS123-C} (IMPACT-Core Predictors)</td>
</tr>
<tr>
<td>AUROC (c-statistic)</td>
<td>0.67</td>
<td>0.67</td>
</tr>
<tr>
<td>Hosmer-Lemeshow p value</td>
<td>0.006</td>
<td>0.03</td>
</tr>
<tr>
<td>Calibration-in-the-large</td>
<td>−0.027</td>
<td>−0.013</td>
</tr>
<tr>
<td>Calibration slope</td>
<td>0.60</td>
<td>0.62</td>
</tr>
<tr>
<td>Brier score</td>
<td>0.23</td>
<td>0.23</td>
</tr>
<tr>
<td>Brier score scaled (%)</td>
<td>4.6</td>
<td>6.6</td>
</tr>
<tr>
<td>Accuracy (%)</td>
<td>65</td>
<td>66</td>
</tr>
<tr>
<td>Sensitivity (%)</td>
<td>72</td>
<td>67</td>
</tr>
<tr>
<td>Specificity (%)</td>
<td>60</td>
<td>66</td>
</tr>
</tbody>
</table>

GOS = Glasgow Outcome Score; IMPACT = International Mission for Prognosis and Clinical Trial design in TBI; CRASH = Corticosteroid Randomisation After Significant Head Injury; LR = logistic regression; AUROC = area under the receiver operating curve.

Median (25–75 percentiles). The last three rows were computed by choosing the cutoff that maximized sensitivity and specificity.

Brier_max = incidence × (1 – incidence) + (1 – incidence) × incidence = 0.25.

Brier score scaled = 1 – (Brier score/Brier_max).

Five patients were excluded as they were younger than 14 yr and therefore out of the calculator’s range.


REFERENCES

ma index that predicts outcome in patients with severe traumatic brain injury. J Trauma 2011; 70:547–553
AUTHOR QUERIES

Author Please Answer All Queries

AQ1—Please confirm if CPP could be expanded as cerebral perfusion pressure.

AQ2—Please check the edits to reference 3.

AQ3—Reference 4 was a duplicate of reference 19 and hence has been deleted and subsequent references renumbered accordingly both in list and text. Please confirm.

AQ4—Please provide the publisher location for reference 19.

AQ5—Please provide the publisher name for reference 20.