Predicting Intracranial Pressure and Brain Tissue Oxygen Crises in Patients with Severe Traumatic Brain Injury

SUPPLEMENTAL MATERIAL

Differences in patient cohorts

Our single-center 6-month mortality for the 3 cohorts, study, model and validation were 28%, 24%, and 14% respectively. These numbers compare favorably with results from recent randomized controlled multi-center trials in severe TBI (EUROTHERM3235 and BEST-TRIP) where mortality among control and intervention groups ranged from 26% to 44%. Our groups are well matched in terms of age, gender and admission Glasgow Coma Scores. We made no attempt and had no intention to create strictly matched groups (that would require propensity score matching) since our primary aim was an ICP/$PbtO_2$ prediction algorithm and not clinical outcome comparisons between groups. Nevertheless, potential explanations for the mortality difference may relate to variability in baseline characteristics (as a cause of trauma more assaults and less transportation related accidents in the older cohort), and to an overall advancement of care leading to decreasing mortality in the more contemporary cohort. Interquartile ranges for number and duration of elevated ICP episodes are comparable among groups and representative of severe TBI populations.

Constructing epochs

Since multiple monitoring devices were sampled, we often had more than one value for a given signal at a given time. Consequently, the first step in preparing our data was to exclude physiologically impossible values and keep

only the first non-excluded data value for each time stamp. We retained values as follows: ICP: 0 - 80; $EtCO_2$: 15 - 80; MAP: 20-160; $PbtO_2$: 0-60; SaO_2: 40-100.

Crises periods were determined by sweeping through the real time data chronologically and tagging periods of time that meet the crisis criteria (ICP or $PbtO_2$ values above or below the crisis threshold for the specified time duration). These periods of time were excluded from the data used to predict crises. To generate the epochs for prediction, we again swept through the data, this time working backwards, starting N minutes before the start of each crisis (to ensure we have some epochs N minutes from crises), extracting all successive 30-minute epochs with at least 25 data points, with no more than 0.05 hours (1.5 minutes) between any two points.

We pre-process each epoch by interpolating any missing points to obtain points at every 0.01 hour (36 seconds), since that was our typical sample frequency. We then downsample the data to every 0.02 hours and smooth the points using a first order Savitzky-Golay filter with window size 3.

We then determined whether the epoch is a precursor to a crisis or not by looking N (15, 30, 60, etc.) minutes in the future and determining if the patient is in a period of time that meets the crisis definition. If so, the epoch is considered pre-crisis, if not, then it is tagged as a non-pre-crisis epoch. If there was a gap in the patient data at the prediction time, the epoch was excluded from consideration for that value of N. All epochs with known status N minutes in the future were used in the AR-OR and logistic regression models for predicting N minutes in the future. The time since last crisis (going back to the start of the

patient stay, if needed) was also used in the AR-OR and some of the logistic regression models.

Tables S1 and S2 show the statistics for the epochs used to predict 30 minutes in the future for the ICP and $PbtO_2$ crises, respectively.

	Study Cohort	Model Selection Cohort	Validation Cohort
Number of epochs	43,353	23,751	38,349
Number of pre-crisis epochs	5,979	3,506	4,025
% of pre-crisis epochs	14%	15%	10%

Table S1: Epoch counts for ICP crises predictions.

	Study Cohort	Model Selection Cohort	Validation Cohort
Number of epochs	7,716	3,303	9,681
Number of pre-crisis epochs	80	30	142
% of pre-crisis epochs	1%	1%	1%

Table S2: Epoch counts for $PbtO_2$ crises predictions.

Tables S3 - S6 show the confusion matrix values for the model selection and validation cohorts. These values are the true positive, false positive, false negative and true negative values for predicting pre-crisis epochs. The values in the tables are based on the "best", non cross-fold validation models as determined by the highest F_2-score. The different F_2-score values were calculated by slowly moving the threshold that divides the pre-crisis label from the non-crisis label and computing the F_2-score after each shift. Using different criteria for "best" will yield different results. For example, if the goal is not to miss any pre-crisis epochs or to minimize false alarms, the threshold selected and the confusion matrix contents would be very different.

		Predicted	
		Pre-crisis	Non-crisis
Actual	Pre-crisis	3,015	491
	Non-crisis	6,382	13,863

Table S3: Confusion matrix for ICP model selection cohort best "AR-OR model + time since last crisis + last 2 ICP values" model (as determined by F_2-score).

		Predicted	
		Pre-crisis	Non-crisis
Actual	Pre-crisis	3,331	694
	Non-crisis	8,327	25,997

Table S4: Confusion matrix for ICP validation cohort best "AR-OR model + time since last crisis + last 2 ICP values" model (as determined by F_2-score).

		Predicted	
		Pre-crisis	Non-crisis
Actual	Pre-crisis	15	15
	Non-crisis	54	3,219

Table S5: Confusion matrix for $PbtO_2$ model selection cohort best "AR-OR model + time since last crisis + last 2 $PbtO_2$ values" model (as determined by F_2-score).

		Predicted	
		Pre-crisis	Non-crisis
Actual	Pre-crisis	100	42
	Non-crisis	293	9,246

Table S6: Confusion matrix for $PbtO_2$ validation cohort best "AR-OR model + time since last crisis + last 2 $PbtO_2$ values" model (as determined by F_2-score).

Additional Model Performance Parameters

Since false positive and false negative values are specific to a model and classification threshold, we provide these values for the optimal model for predicting 30 minutes in the future. We define the optimal model as the model with the highest F_2-score for the rare class. (See the corresponding confusion matrices in Figures S3-S6). Since the prevalence is small for rare class problems, positive and negative likelihood ratios are usually used to measure performance. They are more informative than positive and negative predictive values, where the overall prevalence of the outcomes has a very large impact.

The likelihood ratios and predictive values for the models are provided in the table below. As can be seen, the positive likelihood ratios are well above 1 for the all of the models, indicating that the model prediction provides substantial improvement to the post-test probability of a detection of a crisis, with little change in the post-test probability of no-crisis determination (negative likelihood ratio < 1).

Model	Cohort	Positive Likelihood Ratio	Negative Likelihood Ratio	Positive Predictive Value	Negative Predictive Value
ICP (S3)	Model Selection	2.73	0.20	0.32	0.97
ICP (S4)	Validation	3.41	0.23	0.29	0.97
PbtO2 (S5)	Model Selection	30.31	0.51	0.22	1.00
PbtO2 (S6)	Validation	22.93	0.31	0.25	1.00

Table S7: Positive and Negative Likelihood Ratios and Predictive Values for the best models.

A number of our predictive models utilize a set of features that describe how long it has been since the patient's last crisis.

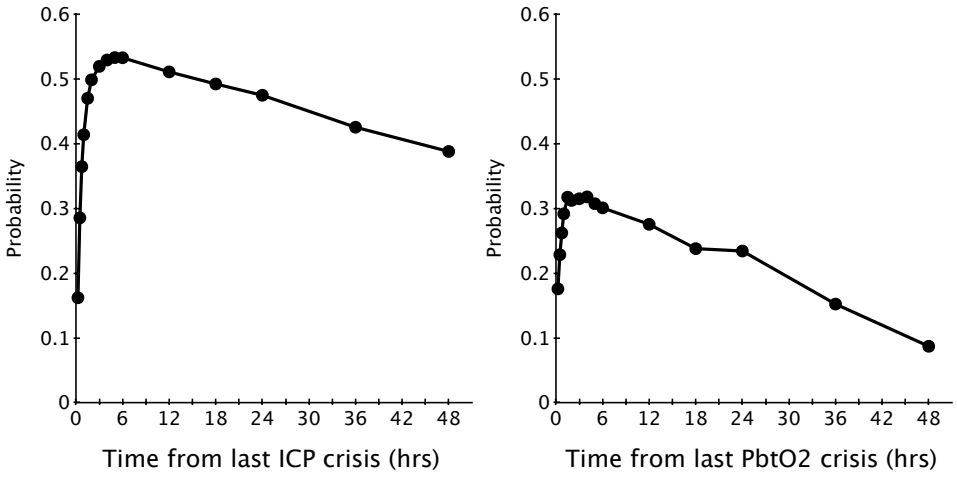

Figure S1: Probability of a second ICP (left) or $PbtO_2$ (right) crisis as a function of time since last crisis.

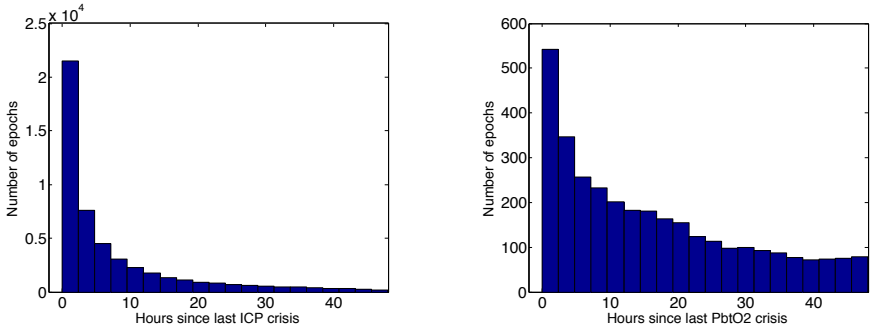

Figure S2: (Left) Histogram of time since last ICP crisis. (Right) Histogram of time since last $PbtO_2$ crisis for training data sets.

Rather than encoding how long it is has been since a crisis using a single number (for example, recording the minutes since last crisis), we bin the different

values into eight categories: (1) ≤ 15 minutes since the last crisis; (2) at least 15 minutes, but ≤ 30 minutes since the last crisis; (3) at least 30 minutes, but ≤ 60 minutes since the last crisis; (4) at least 60 minutes, but ≤ 90 minutes since the last crisis; (5) at least 90 minutes, but ≤ 180 minutes since the last crisis; (6) at least 180 minutes, but ≤ 360 minutes since the last crisis; (7) more than 360 minutes since the last crisis and (8) no prior crisis. We then represent the time since the last crisis with a vector of eight values. Each patient will have a single non-zero value in the vector, with a one indicating which category the patient falls in. Using the different bins as features allows us to model a non-linear relationship with the time since last crisis. If there is no information about the patient crisis state at that time, that is, we have a discontinuity in our data, we exclude that epoch of time from the particular experiment. For reference, Figure S1 shows the likelihood of a second crisis event occurring as a function of time from the first crisis event. Figure S2 depicts a histogram of the time since last crises for all of the 30-minute epochs in the ICP and $PbtO_2$ data sets. In our training data set there were 16,501 epochs without any prior crisis and 7,004 training $PbtO_2$ epochs without any prior crises.

Additional Model Performance Findings (ROC Area):

Time to Predict (min)	Model Selection Cohort							Validation Cohort			
	GP	LR on last 2 ICP values	LR on time since last crisis	LR on last 2 ICP values + time since last crisis	AR-OR model	AR-OR model + time since last crisis	AR-OR model + time since last crisis + last 2 ICP values	AR-OR model + time since last crisis	AR-OR model + time since last crisis, 10-fold cross validation	AR-OR model + time since last crisis + last 2 ICP values	AR-OR model + time since last crisis + last 2 ICP values, 10-fold cross validation
15	-	0.87	0.80	0.88	0.84	0.88	0.90	0.88	0.89	0.89	0.92
30	0.77	0.78	0.80	0.83	0.79	0.84	0.84	0.83	0.85	0.83	0.86
60	-	0.75	0.79	0.81	0.75	0.82	0.82	0.81	0.83	0.81	0.83
120	-	0.73	0.78	0.80	0.74	0.80	0.80	0.79	0.80	0.79	0.81
180	-	0.71	0.76	0.78	0.71	0.78	0.78	0.77	0.78	0.77	0.78
360	-	0.70	0.74	0.76	0.70	0.76	0.76	0.75	0.75	0.75	0.76

Table S8: ROC area for predicting ICP crises for different models and forecasting times.

Time to Predict (min)	Model Selection Cohort							Validation Cohort			
	LR on last 2 $PbtO_2$ values	LR on time since last crisis	LR on last 2 $PbtO_2$ values + time since last crisis	AR-OR model	AR-OR model + time since last crisis	AR-OR model + last 2 PbtO2 values	AR-OR model + time since last crisis + last 2 PbtO2 values	AR-OR model + Last 2 PbtO2 values	AR-OR model + Last 2 PbtO2 values, 10-fold cross validation	AR-OR model + time since last crisis + last 2 PbtO2 values	AR-OR model + time since last crisis + last 2 PbtO2 values, 10-fold cross validation
15	0.98	0.81	0.95	0.92	0.89	0.97	0.94	0.90	0.90	0.90	0.90
30	0.97	0.80	0.91	0.93	0.92	0.97	0.91	0.91	0.90	0.91	0.91
60	0.96	0.87	0.91	0.90	0.89	0.95	0.92	0.83	0.84	0.81	0.83
120	0.88	0.75	0.86	0.84	0.83	0.89	0.86	0.81	0.81	0.78	0.81
180	0.90	0.71	0.85	0.84	0.84	0.89	0.85	0.77	0.77	0.73	0.78
360	0.82	0.62	0.73	0.78	0.72	0.82	0.73	0.72	0.72	0.68	0.74

Table S9: ROC area for predicting $PbtO_2$ crises for different models and forecasting times.

Confidence Intervals

Model	Time to Predict	AUC	CI	
ICPx10	15	.92	0.91	0.92
ICPx10	30	.86	0.85	0.86
ICPx10	60	.83	0.83	0.84
1CPx10	120	.81	0.80	0.81
ICPx10	180	.78	0.78	0.79
ICPx10	360	.76	0.75	0.76
ICP	15	.89	0.88	0.89
ICP	30	.83	0.83	0.84
ICP	60	.81	0.80	0.82
ICP	120	.79	0.78	0.79
ICP	180	.77	0.77	0.78
ICP	360	.75	0.75	0.76
PbtO2x10	15	.90	0.87	0.92
PbtO2x10	30	.91	0.89	0.89
PbtO2x10	60	.83	0.80	0.86
PbtO2x10	120	.81	0.78	0.83
PbtO2x10	180	.78	0.75	0.80
PbtO2x10	360	.74	0.72	0.77
PbtO2	15	.90	0.86	0.92
PbtO2	30	.91	0.87	0.94
PbtO2	60	.81	0.78	0.85
PbtO2	120	.78	0.74	0.81
PbtO2	180	.73	0.70	0.76
PbtO2	360	.68	0.65	0.71

Table S10: The 95% confidence intervals (CI) as a function of prediction time for ICP and $PbtO_2$ crisis events for the best model in tables S8 and S9.

		1 month GOS	3 month GOS	6 month GOS
Signal	**Metric**	**[1,2,3] vs. [4,5]**	**[1,2,3] vs. [4,5]**	**[1,2,3] vs. [4,5]**
ICP	**Minutes in crisis**	**0.104**	**0.001**	**0.003**
	Fraction of time in crisis	**0.501**	**0.200**	**0.049**
$PbtO_2$	**Minutes in crisis**	**0.078**	**0.043**	**0.191**
	Fraction of time in crisis	**0.194**	**0.065**	**0.346**

Table S11: Significance values for correlations in model validation set. Statistically significant values (< 0.05) are shaded.

Predictive Models

Gaussian Process

We follow the process used by Güiza et al. (1) and use the gpml toolbox (http://www.gaussianprocess.org/gpml/) to implement the Gaussian Process predictive model. The features we used are listed in Table S12.

Final Ranking	Predictor Name
1	ICP median over t(-0) - t(-19)
2	ICP minute t(-0)
3	ICP median over t(-0) - t(-9)
4	ICP minute t(-1)
5	ICP minute t(-5)
6	ICP minute t(-4)
7	ICP minute t(-6)
8	ICP median over t(-0) - t(-4)
9	ICP median over t(-5) - t(-9)
10	ICP minute t(-9)
11	ICP minute t(-7)
12	ICP minute t(-3)
13	ICP minute t(-8)
14	ICP median over t(-10) - t(-14)
15	ICP minute t(-11)
16	ICP minute t(-2)
17	ICP minute t(-12)
18	ICP median over t(-10) - t(-19)
19	ICP minute t(-10)
20	ICP minute t(-13)
21	ICP median over t(-15) - t(-19)

22	ICP minute t(-15)
23	ICP minute t(-17)
24	ICP minute t(-16)
25	ICP minute t(-14)
26	ICP minute t(-18)
27	ICP minute t(-19)
28	3rd cepstrum coefficient for MAP
29	2nd cepstrum coefficient for MAP
30	2nd cepstrum coefficient for ICP
31	3rd cepstrum coefficient for ICP
32	Change in ICP from first to last points in epoch
33	ICP standard deviation over t(-0) - t(-9)
34	ICP standard deviation over t(-0) - t(-4)
35	ICP standard deviation over t(-0) - t(-19)
36	Length of Stay
37	5th cepstrum coefficient for ICP
38	4th cepstrum coefficient for ICP
39	CPP minute t(-12)
40	CPP minute t(-13)
41	CPP minute t(-14)
42	CPP median over t(-10) - t(-14)
43	CPP minute t(-4)
44	CPP minute t(-5)
45	CPP minute t(-15)
46	CPP minute t(-19)
47	CPP minute t(-18)
48	CPP minute t(-16)

49	CPP median over t(-0) - t(-9)
50	CPP minute t(-0)
51	CPP minute t(-3)
52	CPP median over t(-15) - t(-19)
53	CPP minute t(-17)
54	CPP minute t(-10)
55	4th cepstrum coefficient for MAP
56	5th cepstrum coefficient for MAP
57	CPP minute t(-6)
58	Largest FFT coefficient for MAP
59	CPP minute t(-9)
60	CPP median over t(-0) - t(-4)
61	CPP median over t(-5) - t(-9)
62	CPP minute t(-1)
63	CPP minute t(-8)
64	CPP minute t(-7)
65	ICP standard deviation over t(-5) - t(-9)
66	CPP minute t(-2)
67	Largest FFT coefficient for ICP
68	ICP standard deviation over t(-15) - t(-19)
69	CPP minute t(-11)
70	4th largest FFT coefficient for ICP
71	5th largest FFT coefficient for ICP
72	MAP median over t(-0) - t(-19)
73	4th largest FFT coefficient for MAP
74	5th largest FFT coefficient for MAP
75	ICP standard deviation over t(-10) - t(-19)

76	CPP median over t(-0) - t(-19)
77	CPP median over t(-10) - t(-19)
78	3rd largest FFT coefficient for MAP
79	CPP standard deviation over t(-0) - t(-4)
80	2nd largest FFT coefficient for MAP
81	MAP minute t(-0)
82	3rd largest FFT coefficient for ICP
83	2nd largest FFT coefficient for ICP
84	ICP standard deviation over t(-10) - t(-14)
85	MAP standard deviation over t(-5) - t(-9)
86	MAP standard deviation over t(-15) - t(-19)
87	change in CPP from first to last points in epoch
88	MAP standard deviation over t(-0) - t(-4)
89	CPP standard deviation over t(-0) - t(-19)
90	MAP standard deviation over t(-10) - t(-14)
91	MAP minute t(-4)
92	MAP standard deviation over t(-0) - t(-9)
93	MAP standard deviation over t(-10) - t(-19)
94	MAP minute t(-6)
95	MAP minute t(-2)
96	MAP minute t(-1)
97	MAP minute t(-12)
98	MAP minute t(-3)
99	MAP minute t(-13)
100	MAP median over t(-0) - t(-9)
101	MAP minute t(-5)
102	MAP minute t(-7)

103	MAP median over t(-0) - t(-4)
104	MAP median over t(-5) - t(-9)
105	MAP median over t(-10) - t(-19)
106	MAP minute t(-15)
107	MAP median over t(-10) - t(-14)
108	MAP minute t(-8)
109	MAP minute t(-14)
110	MAP minute t(-11)
111	MAP minute t(-16)
112	MAP minute t(-9)
113	MAP minute t(-17)
114	MAP median over t(-15) - t(-19)
115	Frequency of the largest FFT coefficient for ICP
116	Frequency of the 2nd largest FFT coefficient for ICP
117	Frequency of the largest FFT coefficient for MAP
118	Frequency of the 2nd largest FFT coefficient for MAP
119	Frequency of the largest FFT coefficient for CPP
120	CPP standard deviation over t(-5) - t(-9)
121	Frequency of the 2nd largest FFT coefficient for CPP
122	2nd cepstrum coefficient for CPP
123	3rd cepstrum coefficient for CPP
124	4th cepstrum coefficient for CPP
125	5th cepstrum coefficient for CPP
126	CPP standard deviation over t(-0) - t(-9)
127	change in MAP from first to last points in epoch
128	change in correlation value
129	MAP minute t(-18)

130	MAP minute t(-10)
131	1st cepstrum coefficient for ICP
132	MAP minute t(-19)
133	correlation between ICP and MAP in over t(-0) - t(-9)
134	5th largest FFT coefficient for CPP
135	1st cepstrum coefficient for MAP
136	3rd largest FFT coefficient for CPP
137	2nd largest FFT coefficient for CPP
138	Correlation between ICP and MAP in t(-10) - t(-19)
139	Largest FFT coefficient for CPP
140	1st cepstrum coefficient for CPP
141	CPP standard deviation over t(-10) - t(-14)
142	4th largest FFT coefficient for CPP
143	CPP standard deviation over t(-10) - t(-19)
144	CPP standard deviation over t(-15) - t(-19)
145	MAP standard deviation over t(-0) - t(-19)

Table S12: Final ranking of epoch features for Gaussian Process experiments

We consistently and randomly select the same number of pre-crisis and non-pre-crisis epochs (982/1964) for study and validation (392/784) to evaluate both the individual features and the top 5%. To test the individual features, we use the 'covRQard' covariance function in the gpml toolbox for 500 iterations. We ranked the features by AUC and selected the top 5% (7 features). We then used the 'covSEiso' covariance function, again with 500 iterations, on the combined features to determine a final AUC for this approach.

Logistic Regression

We use Weka (www.cs.waikato.ac.nz/ml/weka/) version 3.7.12 for logistic regression, and obtained the best results with the default ridge regression value of 1.0E-8.

AR-OR

We augmented the AR-OR model as described in Myers, et al. (2) to weight the points in each epoch in such a way that the points later in the series are given a greater importance. Each point is assigned the weight β^t, where t is the number of minutes from the end of the epoch. Specifically, the last point in the epoch is given weight β^0, or 1, while the first point in the epoch is assigned weight β^{-n}, where n is the number of points in the epoch. The parameter β is learned at the same time as the rest of the model.

Comparing Means using bootstrapping

To compare the median number of crises and time spent in crises between groups of patients with different Glasgow Outcome Scores, we use a null hypothesis of "The median time in crisis for GOS 1-3 is ≤ median time in crisis for GOS 4-5." To test this hypothesis, we followed these steps:

1. From all of the patients, extract total time spent in each type of crisis
2. For each of the GOS months (1, 3, 6)
    a. Remove all data for patients without a valid GOS score
    b. Repeat 10,000 times

i. Sample the data with replacement until we have obtained the same number of sampled patients as original patients after step (i)
ii. Split the values into two groups: 1) GOS 1-3 vs. 2) GOS 4-5
iii. Compute the median value of each group
iv. Compare the medians of each group. If group 1's median is less than or equal to the group 2's median, increment the count for group 1, otherwise, increment the count for group 2.

c. Compute the fraction of time group 1's median is ≤ group 2's. This number is the significance (p-value).

We then repeat this process again using for fraction of time spent in each type of crisis (instead of total time) to check the null hypothesis "The median fraction of time in crisis for GOS 1-3 is ≤ median fraction of time in crisis for GOS 4-5."

REFERENCES

1. Güiza F, Depreitere B, Piper I, Van den Berghe G, Meyfroidt G. 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*. 2013 Feb; 41(2):554-64.
2. Myers RB, Frenzel JC, Ruiz JR, Jermaine CM, Correlating Surgical Vital Sign Quality with 30-Day Outcomes using Regression on Time Series Segment Features, *SIAM International Conference on Data Mining*, 2015.