J Med Syst (2015) 39:14
DOI 10.1007/s10916-014-0187-x

SYSTEMS-LEVEL QUALITY IMPROVEMENT

Predicting Long-Term Outcome After Traumatic Brain Injury Using Repeated Measurements of Glasgow Coma Scale and Data Mining Methods

Hsueh-Yi Lu · Tzu-Chi Li · Yong-Kwang Tu · Jui-Chang Tsai · Hong-Shiee Lai · Lu-Ting Kuo

Received: 1 October 2014 / Accepted: 29 December 2014

Abstract Previous studies have identified some clinical parameters for predicting long-term functional recovery and mortality after traumatic brain injury (TBI). Here, data mining methods were combined with serial Glasgow Coma Scale (GCS) scores and clinical and laboratory parameters to predict 6-month functional outcome and mortality in patients with TBI. Data of consecutive adult patients presenting at a trauma center with moderate-to-severe head injury were retrospectively analyzed. Clinical parameters including serial GCS measurements at emergency department, 7th day, and 14th day and laboratory data were included for analysis (n=115). We employed artificial neural network (ANN), naïve Bayes (NB), decision tree, and logistic regression to predict mortality and functional outcomes at 6 months after TBI. Favorable functional outcome was achieved by 34.8 % of the patients, and overall 6-month mortality was 25.2 %. For 6-month functional outcome prediction, ANN was the best model, with an area under the receiver operating characteristic curve (AUC) of 96.13 %, sensitivity of 83.50 %, and specificity of 89.73 %. The best predictive model for mortality was NB with AUC of 91.14 %, sensitivity of 81.17 %, and specificity of 90.65 %. Sensitivity analysis demonstrated GCS measurements on the 7th and 14th day and difference between emergency room and 14th day GCS score as the most influential attributes both in mortality and functional outcome prediction models. Analysis of serial GCS measurements using data mining methods provided additional predictive information in relation to 6-month mortality and functional outcome in patients with moderate-to-severe TBI.

Keywords Glasgow coma scale · Traumatic brain injury · Mortality · Data mining · Projections and predictions

This article is part of the Topical Collection on *Systems-Level Quality Improvement*

H.-Y. Lu
Department of Industrial Engineering and Management, National Yunlin University of Science and Technology, Douliou City, Yun-Lin county 640, Taiwan
e-mail: hylu@yuntech.edu.tw

T.-C. Li
Department of Nursing, National Taiwan University Hospital, Yun-Lin branch, Yun-Lin county 640, Taiwan
e-mail: ci6964@yahoo.com.tw

Y.-K. Tu · J.-C. Tsai · L.-T. Kuo (✉)
Division of Neurosurgery, Department of Surgery, National Taiwan University Hospital and College of Medicine, No.7, Chungshan S. Rd., Zhongzheng Dist., Taipei City 100, Taiwan
e-mail: kuoluting@gmail.com

Y.-K. Tu
e-mail: yktu@ntu.edu.tw

J.-C. Tsai
e-mail: jctsai@ntu.edu.tw

H.-S. Lai
Department of Surgery, National Taiwan University Hospital, Taipei 100, Taiwan
e-mail: hslai@ntu.edu.tw

Introduction

Traumatic brain injury (TBI), one of the major causes of death and long-lasting disability worldwide, is a disease with heterogeneous mechanisms of injury, pathology, severity, and outcome [1–4]. After acquiring TBI, patients face uncertainty of long-term outcomes regarding behavioral, cognitive, physical and social impairments that could dramatically change their life and productivity [5, 6]. Since the development of intensive care treatments has allowed extremely brain-damaged patients to survive, prediction of the long-term functional outcome becomes a more and more important task.

Published online: 31 January 2015

Identifying a high-risk group for poor outcome in patients after head trauma would allow care and resources to be targeted at those with most to gain and also give the family reasonable expectation to the recovery of neurological function.

Because TBI could lead to multiple organ failures beside brain dysfunction, the development of prediction models is particularly difficult when analyzing the interaction between various clinic attributes to describe a long-term outcome for an individual patient after TBI [7, 8]. To make predictive models useful in practice, reliable, simple, and easy-to-evaluate criteria in a heterogeneous TBI population are required [7, 9–11]. The most common elements constructed in models include sex, age, Glasgow Coma Scale (GCS) score, pupillary reactivity, and laboratory tests [11–13]. GCS is the most widely implemented clinical scoring system to assess level of consciousness after TBI, which classifies the severity as mild (GCS 13–15), moderate (GCS 9–12), or severe (GCS < 9) [14–16]. Studies have demonstrated that GCS, which is easy to assess and reproducible, is an important prognostic factor in outcome prediction models [16–19] and may influence treatment strategies [20, 21]. However, how changes in GCS scores at various time points after hospital admission relate to outcome is unknown. Previous reported prognostic models were based on multiple logistic regression, which mainly explored the significant relationships among the variables in a group of individuals and have limitations when making decision at the individual patient level [17, 22–24].

Data mining is a computational process for extracting previously unknown and potentially useful information from data using a combination of artificial intelligence, machine learning, statistics, and database systems [25–27]. This knowledge discovery methodology has been increasingly used in medical studies to explore the patterns or relationships among medical variables and make predictions about unknown values at the individual level by mining historical medical data [28–31]. Some predictive data mining methods have been introduced to predict prognostic outcomes of patients with TBI and are potentially more accurate than conventional statistical methods [7, 9, 32, 33]. Most of studies used the predictors at the time of admission; only one study by Lee and colleagues measured GCS at multiple time points and reported that patients with low GCS at 24 and 48 h after TBI had unfavorable outcome [34].

The objective of this study was to use data mining methodologies that combine serial assessments of GCS scores as predictors for 6-month outcomes in patients with moderate-to-severe TBI. We hypothesized that serial GCS measurements could substantially increase the predictive ability to discriminate TBI outcomes. Four predictive models were developed and compared based on a limited set of variables that are rapidly and easily assessable in routine practice. We anticipate that this approach, with its ease and simplicity, will enable reliable prognostic prediction.

Material and methods

Patients

In this retrospective study, the study samples consisted of 128 consecutive patients admitted to the neurosurgical intensive care unit due to TBI at the National Taiwan University Hospital Yun-Lin Branch from 2009 to 2012. We included only patients aged 18 years or older with moderate-to-severe brain injury, and excluded patients who died within 14 days after admission or with incomplete data. TBI included the following diagnoses due to head injury: cerebral contusion, subdural hematoma, epidural hematoma, subarachnoid hemorrhage, and diffuse axonal injury. Only patients with complete documents of GCS scores (upon hospital arrival, 7th day, and 14th day after head injury) and 6-month outcome were included in the study. This study was approved by National Taiwan University Hospital's Institutional Review Board (IRB case #201304090RINC).

On arrival to ER, the patients' clinical, and lab test data were collected according to the Advanced Trauma Life Support guidelines and the American Association of Neurological Surgeons/Congress of Neurological Surgeons Guidelines for the Management of Head Injury. On admission, the GCS was evaluated by physicians at emergency room to describe the initial neurological status. The decision to operate on an acute subdural hematoma, epidural hematoma or other intracranial hemorrhages with decompressive surgery was based on the patient's GCS score, pupillary exam, comorbidities, CT findings, age, and whether there was neurological deterioration. Post-operative medical management included ventilation, oxygenation, head elevation (30°), fluid resuscitation, and sedation with either sedatives.

Outcome measure

Outcome was assessed by neurosurgeons in the outpatient clinic using the 6-month Glasgow Outcome Scale (GOS) score [35]. The GOS is a 5-point scale dividing patients into five categories: 1, dead; 2, vegetative state; 3, severe disability; 4, moderate disability; and 5, good recovery [36]. In this study, we developed two types of prediction models to predict 6-month functional outcome and mortality. For the prediction of functional outcome, the outcome variable was coded into a binary system of unfavorable outcome (GOS 1–3) or favorable outcome (GOS 4–5) [7, 22–24]. For the mortality prediction, the outcome variable was dichotomized into dead (GOS 1) or alive (GOS 2–5).

Predictors

Since we were trying to develop and compare predictive models mainly based on easy-to-assess and routinely collected

data, the predictors included demographic, clinical, and lab test data collected from emergency room (ER) and intensive care unit (ICU) admission, which included GCS, age, gender, blood glucose level, white blood cell, diabetes mellitus history, hypertension history, pupil size, and diagnosis (Table 1). These predictors have been reported to be associated with morbidity or mortality for patients with TBI [35, 37–41] and were coded into nominal or ordinal scales in order to render the input data suitable for processing models. To dichotomizing the continuous variables (age, blood glucose level, and white bold cell), this study adopted a outcome-based methods which iterates all observed values for the one which best separates the risk groups with respect to the study outcomes according to a chi-squared test [42]. The initial GCS score was measured upon emergency room (ER) arrival. Another two time points of GCS measurements (7th day and 14th day) were included for analysis. The changes of GCS scores from ER to the 7th day, from ER to the 14th day, and from the 7th to 14th day were calculated and included for the analysis. Through a pilot process, an inter-reader reliability [43] was conducted (ICC $_{(2,\ 1)}$ coefficients=0.94) to ensure that the GCS scores were estimated reliably between medical staff.

Data mining

This study employed three data mining methods, artificial neural network (ANN), naïve Bayes (NB), and decision tree (DT), and one statistical method, logistic regression (LR), to classify patients' favorable/unfavorable and alive/dead outcomes at 6 months after TBI. To run those prediction models, we used the open source WEKA software package (http://www.cs.waikato.ac.nz/ml/weka/, version 3.6.6) created at the University of Waikato in New Zealand.

ANN

The ANN is a mathematical model that mimics the human brain's own decision process to employ a nonlinear structure with previously solved examples and build a system of neurons to make new predictions [44]. The advantages of using an ANN is that it could process nonlinear relationships among dependent and independent variables whose relationships are complex, multidimensional, and interactive [9, 33, 45].

Naive Bayes

Naïve Bayes is one of the simplest supervised learning algorithms for data mining, which applies Bayes' theorem with the assumption of conditional independence of features regarding class to estimate a high-dimensional probability by reducing it to one-dimensional conditional probability [25]. In reality, this assumption may not be true, but it could dramatically simplify the complexity of the model and efficiently process the classification task.

Table 1 Variable list

Variables	Type	Coding
Dependent		
GOS	Ordinal	Range 1~5
6-month functional outcome	Nominal	2 codes (0 for GOS 1–3 as unfavorable, 1 for GOS 4–5 as favorable)
6-month mortality	Nominal	2 codes (0 for alive, 1 for dead)
Independent		
Age	Ordinal	2 codes (0 for ≤ 44, 1 for > 44)
Gender	Nominal	2 codes (0 for female, 1 for male)
Blood glucose level in ER	Ordinal	2 codes (0 for ≤ 190, 1 for > 190)
White blood cell (10^9/L) in ER	Ordinal	2 codes (0 for ≤ 16, 1 for > 16)
Diabetes mellitus history	Nominal	2 codes (0 for no, 1 for yes)
Hypertension history	Nominal	2 codes (0 for no, 1 for yes)
Pupil size	Nominal	2 codes (0 for normal, 1 for one or two dilated)
Diagnosis	Nominal	3 codes (2 for *Subdural hematoma*, 3 for *epidural hematoma*, 4 for *others*)
GCS in ER	Ordinal	Range 3~12
GCS 7th day	Ordinal	Range 3~15
GCS on 14th day	Ordinal	Range 3~15
GCS change from ER to 7th day[a]	Ordinal	Range −12~12
GCS change from ER to 14th day	Ordinal	Range −12~12
GCS change from 7th to 14th day	Ordinal	Range −12~12

ER emergency room, *GOS* Glasgow Outcome Scale, *GCS* Glasgow outcome scale

[a] GCS change: subtract later measured GCS from earlier one

Decision tree

A decision tree is a tree-like graph produced by repeated partitioning of the dataset, maximizing data separation into more homogeneous subgroups to display decisions and possible outcomes [46, 47]. A decision tree can incorporate nonlinear relationships and interactions of classifications with a simple representation of the data, making interpretation and determination of rules very easy [48]. Its effectiveness in many well-developed classification algorithms, such as ID3, C4.5, C5 [46, 48, 49] and CART [47], has resulted in its widespread use in medical research [48]. In this study, we chose to use the C4.5 algorithm as our decision tree method.

K-fold cross-validation

To minimize the generalization error associated with randomness leading to biased estimation of future examples, K-fold cross-validation is a commonly computational technique to validate how well prediction models generalize to unseen data [50, 51]. It shows the ability to lower the prediction variance and avoid the bias of over-fitting on the training data [52, 53]. A 10-fold cross-validation was used in our study due to the efficiency to complete the models [54, 55].

Performance measurement

To evaluate the performance of our prediction models, calibration and discrimination were both measured. Calibration (sometimes named, reliability) for discrete classification is defined as the degree of agreement between predicted and observed outcomes. In our binary prediction models, it can be expressed as a match between the expected value of the proportion of classes and the observed value. For instance, if our data has a distribution of 65 % of unfavorable cases and 35 % of favorable cases, a model predicting 95 % of unfavorable cases and 5 % of favorable cases is not well calibrated. Two statistics, γ^2 and correct rate, were used to measure the calibration of the predictive models. γ^2 was computed as the amount of the true class distribution with the estimated one. The smaller the values, the better the calibration, with a perfectly calibrated model having a value of zero [56]. The correction rate ([true positive + true negative]/total) was measured to determine the global fit or calibration accuracy and 100 % correction rate represents as a perfect calibration.

Discrimination (sometimes named resolution) refers to the ability of a predictive model to distinguish patients with one outcome from those with another. In this study, sensitivity, specificity, and the area under the receiver operating characteristic curve (AUC) were adopted to analyze discrimination power of the models. For each prediction model, an AUC plot were constructed to iterate all possible cut-off points to identify a best representation of the trade-off between sensitivity and specificity [57]. A perfect prediction model would yield a 100 % sensitivity and 100 % specificity [58]. The value of AUC varies from 0.5 (performs no better than chance alone) to 1.0 (perfectly separates two populations) [57]. The AUC is independent of both the distribution of the class and of any threshold cutoff used to classify outcomes [59].

In each prediction model, 100 experiments with a 10-fold cross-validation approach were conducted to minimize the bias associated with random sampling of training/test datasets. Performance measures were calculated and compared using paired t-tests ($P<0.05$) for every pairs between four approaches (ANN, NB, DT and LR) of the prediction models (100 experiment results in each model).

Sensitivity analysis for important attributes

The importance of attributes was evaluated by a sensitivity analysis to measure the contribution of each attribute to the discriminatory abilities of the prediction models. The analysis started with a series of leave-out comparisons to compare the baseline AUC (all attributes in the model) with the AUC reassessed by leaving one attribute alone in the model. For each attribute and all algorithms, we ran baseline AUC and leave-one-alone AUC 100 times each, and then compared these AUCs with a t-test.

Results

Of the 128 patients with TBI who were treated during the study period, 10 patients died within 14 days after admission and three patients had incomplete data. Finally, 115 patients (89.8 %) were include in this study. The baseline demographic and clinic variables are outlined in Table 2. The mean age was 55.47 years, blood glucose level was 161.10, and white blood cell (10^9/L) was 12.94. Overall, 68.7 % of the patients were male, 15.7 % had diabetes mellitus, 28.7 % had hypertension history and 86.1 % had normal pupil size. The GCS scores in ER, on the 7th day and on the 14th day were mostly distributed between 3 and 8 (70.4 %, 52.2 % and 42.6 % respectively). Changes of GCS scores from ER to the 7th day, from ER to the 14th day, and from the 7th to 14th day were mostly distributed at 2~−2 (49.6 %), more than 2 (45.2 %) and 2~−2 (85.2 %), respectively.

Prediction of functional outcome

Favorable functional recovery was achieved by 40 (34.8 %) patients in our series (Table 2). Patients with unfavorable outcomes were significantly older (60.88 years for unfavorable outcomes and 45.33 years for favorable outcomes) (Table 3). Predominant proportion of patients with unfavorable

Table 2 Distributions of variables

Description	Mean/Frequency	S.D./Percentage
Age	55.47	19.09
Blood glucose level in ER	161.10	72.01
White blood cell (10^9/L) in ER	12.94	5.87
GCS in ER (*3–8/9–12*)	81/34	70.4/29.6
GCS on 7th day (*3–8/9–12/13–15*)	60/28/27	52.2/24.3/23.5
GCS on 14th day (*3–8/9–12/13–15*)	49/30/36	42.6/26.1/31.3
GCS change from ER to 7th day (*>2, 2∼−2, < −2*)[a]	43/57/15	37.4/49.6/13.0
GCS change from ER to 14th day (*>2, 2∼−2, < −2*)	52/49/14	45.2/42.6/12.2
GCS change from 7th to 14th day (*>2, 2∼−2, < −2*)	13/98/4	11.3/85.2/3.5
Gender (*female/male*)	36/79	31.3/68.7
Diabetes mellitus history (*yes/no*)	18/97	15.7/84.3
Hypertension history (*yes/no*)	33/82	28.7/71.3
Pupil size (*normal/one or two dilated*)	99/16	86.1/13.9
Diagnosis (*subdural hematoma/ epidural hematoma/others*)	90/13/12	78.3/11.3/10.4
6-mon functional outcome (*unfavorable/favorable*)	75/40	65.2/34.8
6-mon mortality (*alive/dead*)	86/29	74.8/25.2

ER emergency room, *WBC* white blood cell, *GCS* Glasgow outcome scale

[a] GCS change: subtract later measured GCS from earlier one

outcomes had lower GCS scores at all time points (ER, 7th day, and 14th day) and significantly slower improvement in GCS scores over these time points.

Correct rate and γ^2 suggested that the models in predicting 6-month functional outcome were essentially calibrated well (Table 4). The best prediction model was ANN, which had better calibration in term of γ^2 (0.061), significantly lower than NB and DT (0.665 and 0.147, respectively). ANN had the significantly highest performance regarding AUC (96.13 %) with 83.50 % sensitivity and 89.73 % specificity. The results are noteworthy in that all of our prediction models were able to discriminate functional outcome with excellent AUCs (NB: 94.45 %, DT: 91.86 %, LR: 92.47 %).

Table 3 Characteristics of the patients categorized as dichotomous 6-month outcomes

Variable	Favorable (n=40)	Unfavorable (n=75)	p^*	Alive (n=86)	Dead (n=29)	p
Age (years)	45.33 (19.67)[a]	60.88 (16.51)	0.000	54.10 (19.50)	59.52 (17.52)	0.188
Blood glucose level in ER	149.98 (80.45)	167.03 (66.89)	0.228	159.69 (76.46)	165.31 (57.77)	0.717
WBC (10^9/L) in ER	13.19 (5.09)	12.81 (6.27)	0.749	12.51 (4.70)	14.23 (8.41)	0.174
GCS in ER (3–8/9–12)	22/18	59/16	0.008	56/30	25/4	0.031
GCS on 7th day (3–8/9–12/13–15)	4/14/22	56/14/5	0.000	33/27/26	27/1/1	0.000
GCS on 14th day (3–8/9–12/13–15)	0/11/29	49/19/7	0.000	22/29/35	27/1/1	0.000
GCS change from ER to 7th day (>2, 2∼−2, < −2)[b]	27/12/1	16/45/14	0.000	39/39/8	4/18/7	0.005
GCS change from ER to 14th day (>2, 2∼−2, < −2)	31/8/1	21/41/13	0.000	47/33/6	5/16/8	0.000
GCS change from 7th to 14th day (>2, 2∼−2, < −2)	8/32/0	5/66/4	0.040	12/72/2	1/26/2	0.173
Gender (female/male)	10/30	26/49	0.287	28/58	8/21	0.618
Diabetes mellitus history (no/yes)	34/6	63/12	0.888	70/16	27/2	0.133
Hypertension history (no/yes)	32 /8	50/25	0.132	62/24	20/9	0.747
Pupil size (normal/one or two dilated)	35/5	64/11	0.749	75/11	24/5	0.549
Diagnosis (subdural hematoma/ epidural hematoma/ others)	29/10/1	61/3/11	0.001	67/13/6	23/0/6	0.016

ER emergency room, *WBC* white blood cell, *GCS* Glasgow outcome scale

[*] p-values were accessed to examine the homogeneity within favorable/unfavorable groups by t-test for continuous variables and chi-square test for categorical variables

[a] Standard deviation

[b] GCS change: subtract later measured GCS from earlier one

Table 4 Prediction performances

Model	Calibration		Discrimination				
	Correction rate	γ^2	Sensitivity	Specificity	AUC	LR+ [a]	LR- [a]
Functional outcome prediction							
ANN	87.54	0.061	83.50	89.73	96.13	8.13	0.18
NB	89.40*	0.665*	92.25*	87.90*	94.45*	7.62	0.09
DT	87.58	0.147*	80.83*	91.21*	91.86*	9.20	0.21
LR	86.91	0.056	81.08*	90.11	92.47*	8.20	0.21
Mortality prediction							
NB	88.23	0.165	81.17	90.65	90.14	8.68	0.21
ANN	82.67***	0.259	61.56***	89.77	81.04***	6.02	0.43
DT	86.78***	0.130	70.00***	92.31***	77.85***	9.10	0.32
LR	85.39***	0.132	68.44***	90.99	87.29***	7.60	0.35

AUC area under the receiver operating characteristic curve, *ANN* artificial neural network, *NB* naïve Bayes, *DT* decision tree, *LR* logistic regression

* Statistically significant (Wilcoxon test where $p<0.05$) difference comparing to ANN model

*** Statistically significant (Wilcoxon test where $p<0.05$) difference comparing to NB model

[a] LR+, LR−: likelihood ratios for positive and negative results

Prediction of mortality

The distribution of mortality was 86 and 29 patients for "alive" and "dead" with a mortality rate of 25.2 % (Table 2). The same GCS pattern was found in mortality prediction in that the GCS scores at all measurement points (ER, 7th day, and 14th day) were generally lower in the group of patients with mortality (Table 3). The patients who were alive had significantly better GCS improvements from ER to 7th day and from ER to 14th day.

In general, the models were calibrated well in term of correct rate and γ^2, particularly for NB, which had a significantly better correct rate than other three models (Table 4). The best predictive model for mortality was NB with prediction performances including 90.14 % AUC. The DT model had a relatively lower performance (77.85 % AUC) compared to the other three models.

Important attributes

A sensitivity analysis was conducted to identify the importance that each attribute in the prediction models. In Table 5, six attributes with the most importance are listed in order of the averaged negative logarithms of test P-value, where the smaller negative logarithms (the larger P-values) were less changes between baseline and leave-one-alone AUC and represent greater contributions to the prediction models. As anticipated, most of the influential attributes were GCS scores measured on the 7th day, on the 14th day, and the GCS change from ER to the 14th day.

Discussion

The comparison analysis showed that our four prediction models, which applied three data mining algorithms and one statistical approach, can predict functional outcome and

Table 5 Attribute importance

Attribute (top six)	-log (P value)*				Average
	ANN	NB	DT	LR	
Function outcome prediction					
GCS on 7th day	7.90	8.37	7.34	0.92	5.54
GCS on 14th day	2.94	8.10	7.61	1.04	5.58
GCS change from ER to 14th day	17.44	14.58	11.93	9.81	12.11
GCS change from ER to 7th day	23.12	21.10	28.53	16.56	22.06
Age	29.80	30.25	22.54	24.76	25.85
GCS in ER	28.24	30.99	45.27	20.56	32.27
Mortality prediction					
GCS on 7th day	6.15	0.22	2.19	0.89	2.36
GCS on 14th day	9.49	1.76	6.24	3.67	5.29
GCS change from ER to 14th day	0.41	6.52	16.92	4.63	7.12
GCS in ER	3.34	14.41	2.38	8.56	7.17
GCS change from ER to 7th day	3.87	14.47	27.14	9.72	13.80
Age	16.99	25.43	27.14	25.21	23.69

ER emergency room, *GCS* Glasgow outcome scale, *ANN* artificial neural network, *NB* naïve Bayes, *DT* decision tree, *LR* logistic regression

* Negative logarithm of p-value obtained from T-test

mortality for patients with TBI. Nevertheless, data mining models provided the best fit. We propose that the ANN model with 96.13 % AUC is optimal for predicting 6-month functional outcome and the NB with 90.14 % AUC is the best algorithm for predicting 6-month mortality. Several studies have evaluated the ability of prediction models to correctly predict outcomes for patients with TBI and showed a wide range of AUCs (65 to 85 %) [17, 18, 22]. We found that our data mining models (ANN, NB, DT) accurately discriminated functional outcome with an AUC of 92–96 % and mortality with an AUC of 78–91 %, which compared favorably with AUCs reported in previous studies.

Age, GCS scores, pupillary reaction, and image results are significant predictors of outcomes in patients with TBI [60]. Findings in our study are consistent with previous research on age and GCS scores. The results from the analysis of important attributes (Table 5) indicate that the GCS scores at different time points (ER, 7th, and 14th day), GCS changes between time points (from ER to 7th day and from ER to 14th day), and age were the most influential variables in the performance of our prediction models. As the data were collected in a tertiary trauma center in an agricultural county where persons 65 years or older represent more than 14 % of the population and motorcycle is the dominant method of transportation, the mean age of these patients with head injury, 55.5 years, is older than in most previous studies of adults with TBI [9, 18, 61–63]. Pupillary size was shown to be a non-significant predictor in our study. Anisocoria is commonly detected in patients with glaucoma or after eye surgery, and the prevalence of glaucoma is age-specific and higher in the older people. Therefore, pupillary size has some limitations to be correctly analyzed in some of these patients. Excluding the patients who died within 14 days since admission, our 6-month unfavorable functional recovery and mortality rate of 65.2 and 25.2 % are comparable to previous studies [11, 18, 22, 32, 61, 63] , which reported unfavorable functional recovery rates (GOS 1–3) of 50–60 % and mortality rates of 20–30 %.

Although GCS often varies dynamically during the early stage of the injury, we recorded and analyzed GCS scores upon admission. It is still unclear which time points of GCS measurement have better prognostic value. A number of studies have described the relationship between GCS scores at a certain time point after TBI and the prognosis for patients with head injuries [63–65]. Those studies used GCS at scene, post-resuscitation GCS, or GCS at ICU discharge and found their usefulness in outcome prediction. In this study, we used serial GCS measurements upon admission, 7th day, and 14th day in patients after TBI to help predict long-term functional

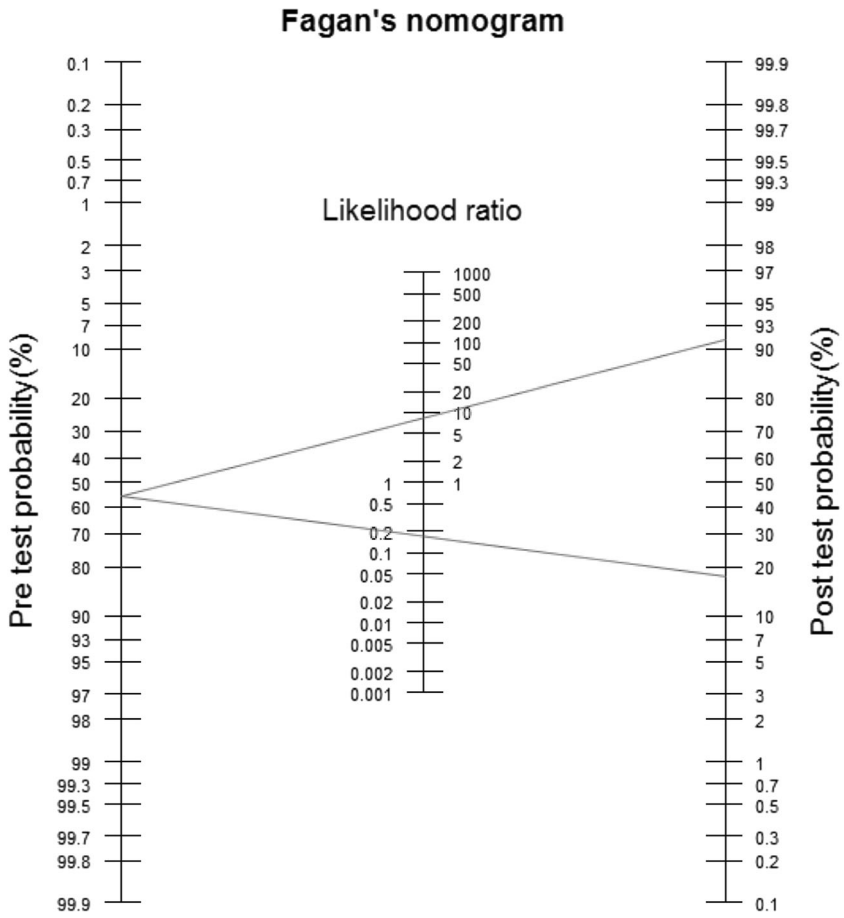

Fig. 1 The use of the Fagan's nomogram (a straight line through the pretest probability of 55 % and the LR+ of 8.13 yields a posttest probability of 91 %; a straight line through the pretest probability of 55 % and the LR- of 0.18 yields a posttest probability of 18 %)

outcome and mortality. Our analysis led to a better predictive value than using initial GCS alone or measurements at another single time point. To our knowledge, this is the first study to employ multiple GCS measurements and their interval changes in prediction models. In our study, sensitivity analysis demonstrates the superiority of serial GCS (on 7th day, 14th day, and the GCS change from admission to 14th day) over other clinical features (blood glucose level, white blood cell, diabetes mellitus history, hypertension history, and diagnosis) in outcome prediction (Table 5). In contrast to the complex analysis of imaging findings or various clinical data, we attempted to use practical and easily assessable measurements to establish reliable prediction models.

Current outcome prediction models in TBI generally express output as a dichotomous endpoint; however, in terms of medical decision support, it is more informative to provide physicians the likelihood of a specific outcome with a degree of confidence that the patient is correctly classified into one group. Therefore, to assess the probability of outcomes for an individual patient, we propose combining our prediction results with likelihood ratios and Fagan's nomogram. Both positive likelihood ratio (LR+) and negative likelihood ratio (LR–), representing the ratio of positive and negative classes of binary prediction results, are calculated by using sensitivity and specificity values from a prediction model [66, 67]. As shown in Table 4, the sensitivity and specificity of the ANN model were 83.5 and 89.73 %, respectively, which yields a LR+ of 8.13 and a LR– of 0.18. In practice, neurosurgeons have an initial judgment on the likelihood of a possible outcome, referred to as a pretest probability, based on their knowledge and experience. However, an intermediate pretest probability can be vague, which introduces certain uncertainty. Based on Bayes' theorem, this pretest probability could be refined to yield a posttest probability by using likelihood ratios and a prediction result as an attempt to supplement the subjective clinical judgment with consideration of objective evidence-based estimates [66, 68]. Fagan's nomogram is a useful graphical tool to easily assess the posttest probability once the pretest probability and likelihood ratios of a prediction result are known [69]. For example, if the initial judgment of an unfavorable outcome for a patient is 55 %, and our ANN model classified this patient into a "unfavorable" outcome with the LR+ estimated at 8.13 (Table 4), a straight line (Figure 1) drawn from the pretest probability of 55 % through the LR+ of 8.13 intersects with the posttest probability of approximately 91 %. The probability of having an unfavorable outcome increases from 55 to 91 %. Alternately, when the prediction result is "favorable," the probability of an unfavorable outcome drops from 55 to 18 %. By using this nomogram, our predictive data mining results could be translated into a probability of an outcome for individual patients simply and easily, which is important for predicting prognosis. In a clinic setting, Fagan's nomogram could be embedded as a tool in hospital information systems to take the advantage of information technology to carry out the calculations and obtain the graphics.

Some limitations in our study should be addressed in future research. This study was conducted in a single institution, and our study needs external validation to get an accurate measure of performance outside the studied population to confirm its generalizability. Other data mining methods such as support vector machines or Bayesian networks could be developed to explore improvements of the prediction performances. We did not include pupillary size, one of the significant signs related to TBI prognosis, in the prediction models due to non-significance to the prediction results. Although assessing serial GCS in patients with TBI in the ER and ICU is common, the accuracy of GCS measurements may be limited by insufficient inter-rater reliability.

In conclusion, this study used a series of repeated assessments of GCS scores to predict functional outcome and mortality of patients with TBI after 6 months. We developed three data mining models (ANN, NB, and DT) and compared them with a statistical model (LR) based on early clinical and laboratory data after TBI. Since our study is aimed to enhance and not replace a physician's clinical judgment, combined with likelihood ratios and Bayes' theorem, our models provide an additional tool that is simple and affordable for neurosurgeons to classify patients with TBI as well as estimate the probability of favorable functional outcome and mortality to assist prognostic decision-making.

References

1. Traumatic brain injury: Time to end the silence. *Lancet Neurol*. 9(4): 331, 2010. doi:10.1016/S1474-4422(10)70069-7.
2. Ghajar, J., Traumatic brain injury. *Lancet* 356(9233):923–929, 2000. doi:10.1016/S0140-6736(00)02689-1.
3. Maas, A. I., Stocchetti, N., and Bullock, R., Moderate and severe traumatic brain injury in adults. *Lancet Neurol*. 7(8):728–741, 2008. doi:10.1016/S1474-4422(08)70164-9.
4. Roozenbeek, B., Chiu, Y. L., Lingsma, H. F., Gerber, L. M., Steyerberg, E. W., Ghajar, J., and Maas, A. I., Predicting 14-day mortality after severe traumatic brain injury: Application of the IMPACT models in the brain trauma foundation TBI-trac(R) New York State database. *J. Neurotrauma* 29(7):1306–1312, 2012. doi: 10.1089/neu.2011.1988.
5. Perez, R., Costa, U., Torrent, M., Solana, J., Opisso, E., Caceres, C., Tormos, J. M., Medina, J., and Gomez, E. J., Upper limb portable motion analysis system based on inertial technology for neurorehabilitation purposes. *Sensors (Basel)* 10(12):10733–10751, 2010. doi:10.3390/s101210733.
6. Silver, J. M., McAllister, T. W., and Yudofsky, S. C., *Textbook of traumatic brain injury*, 2nd edition. American Psychiatric Pub, Washington, DC, 2011.
7. Brown, A. W., Malec, J. F., McClelland, R. L., Diehl, N. N., Englander, J., and Cifu, D. X., Clinical elements that predict outcome

after traumatic brain injury: A prospective multicenter recursive partitioning (decision-tree) analysis. *J. Neurotrauma* 22(10):1040–1051, 2005. doi:10.1089/neu.2005.22.1040.

8. Lingsma, H. F., Roozenbeek, B., Steyerberg, E. W., Murray, G. D., and Maas, A. I., Early prognosis in traumatic brain injury: From prophecies to predictions. *Lancet Neurol.* 9(5):543–554, 2010. doi:10.1016/S1474-4422(10)70065-X.
9. Eftekhar, B., Mohammad, K., Ardebili, H. E., Ghodsi, M., and Ketabchi, E., Comparison of artificial neural network and logistic regression models for prediction of mortality in head trauma based on initial clinical data. *BMC Med. Inf. Decis. Making* 5:3, 2005. doi:10.1186/1472-6947-5-3.
10. Wyatt, J. C., and Altman, D. G., Prognostic models: Clinically useful or quickly forgotten? *Brit. M. J.* 311:1539–1541, 1995.
11. Rovlias, A., and Kotsou, S., Classification and regression tree for prediction of outcome after severe head injury using simple clinical and laboratory variables. *J. Neurotrauma* 21(7):886–893, 2004. doi:10.1089/0897715041526249.
12. Quigley, M. R., Vidovich, D., Cantella, D., Wilberger, J. E., Maroon, J. C., and Diamond, D., Defining the limits of survivorship after very severe head injury. *J. Trauma* 42(1):7–10, 1997.
13. Schreiber, M. A., Aoki, N., Scott, B. G., and Beck, J. R., Determinants of mortality in patients with severe blunt head injury. *Arch. Surg.* 137(3):285–290, 2002.
14. Rimel, R. W., Jane, J. A., and Edlich, R. F., An injury severity scale for comprehensive management of central nervous system trauma. *JACEP* 8(2):64–67, 1979.
15. Knaus, W. A., Draper, E. A., Wagner, D. P., and Zimmerman, J. E., APACHE II: A severity of disease classification system. *Crit. Care Med.* 13(10):818–829, 1985.
16. Teasdale, G., and Jennett, B., Assessment of coma and impaired consciousness. A practical scale. *Lancet* 2(7872):81–84, 1974.
17. Collaborators, M. C. T., Perel, P., Arango, M., Clayton, T., Edwards, P., Komolafe, E., Poccock, S., Roberts, I., Shakur, H., Steyerberg, E., and Yutthakasemsunt, S., Predicting outcome after traumatic brain injury: Practical prognostic models based on large cohort of international patients. *BMJ* 336(7641):425–429, 2008. doi:10.1136/bmj.39461.643438.25.
18. Roozenbeek, B., Lingsma, H. F., Lecky, F. E., Lu, J., Weir, J., Butcher, I., McHugh, G. S., Murray, G. D., Perel, P., Maas, A. I., Steyerberg, E. W., and International Mission on Prognosis Analysis of Clinical Trials in Traumatic Brain Injury Study G, Corticosteroid Randomisation After Significant Head Injury Trial C, Trauma A, Research N, Prediction of outcome after moderate and severe traumatic brain injury: External validation of the International Mission on Prognosis and Analysis of Clinical Trials (IMPACT) and Corticoid Randomisation after Significant Head injury (CRASH) prognostic models. *Crit. Care Med.* 40(5):1609–1617, 2012. doi:10.1097/CCM.0b013e31824519ce.
19. Duncan, R., and Thakore, S., Decreased Glasgow Coma Scale score does not mandate endotracheal intubation in the emergency department. *J. Emerg. Med.* 37(4):451–455, 2009. doi:10.1016/j.jemermed.2008.11.026.
20. Jain, S., Dharap, S. B., and Gore, M. A., Early prediction of outcome in very severe closed head injury. *Injury* 39(5):598–603, 2008. doi:10.1016/j.injury.2007.06.003.
21. Marmarou, A., Lu, J., Butcher, I., McHugh, G. S., Murray, G. D., Steyerberg, E. W., Mushkudiani, N. A., Choi, S., and Maas, A. I., Prognostic value of the Glasgow Coma Scale and pupil reactivity in traumatic brain injury assessed pre-hospital and on enrollment: An IMPACT analysis. *J. Neurotrauma* 24(2):270–280, 2007. doi:10.1089/neu.2006.0029.
22. Steyerberg, E. W., Mushkudiani, N., Perel, P., Butcher, I., Lu, J., McHugh, G. S., Murray, G. D., Marmarou, A., Roberts, I., Habbema, J. D., and Maas, A. I., Predicting outcome after traumatic brain injury: Development and international validation of prognostic scores based on admission characteristics. *PLoS Med.* 5(8):e165, 2008. doi:10.1371/journal.pmed.0050165. discussion e165.
23. Stevens, R. D., and Sutter, R., Prognosis in severe brain injury. *Crit. Care Med.* 41(4):1104–1123, 2013. doi:10.1097/CCM.0b013e318287ee79.
24. Galanaud, D., Perlbarg, V., Gupta, R., Stevens, R. D., Sanchez, P., Tollard, E., de Champfleur, N. M., Dinkel, J., Faivre, S., Soto-Ares, G., Veber, B., Cottenceau, V., Masson, F., Tourdias, T., Andre, E., Audibert, G., Schmitt, E., Ibarrola, D., Dailler, F., Vanhaudenhuyse, A., Tshibanda, L., Payen, J. F., Le Bas, J. F., Krainik, A., Bruder, N., Girard, N., Laureys, S., Benali, H., Puybasset, L., and Neuro Imaging for Coma E, Recovery C, Assessment of white matter injury and outcome in severe brain trauma: A prospective multicenter cohort. *Anesthesiology* 117(6):1300–1310, 2012. doi:10.1097/ALN.0b013e3182755558.
25. Han, J., and Kamber, M., *Data mining: concepts and techniques*. Diane Cerra, San Francisco, CA, 2006.
26. Bellazzi, R., and Zupan, B., Predictive data mining in clinical medicine: Current issues and guidelines. *Int. J. Med. Inform.* 77(2):81–97, 2008. doi:10.1016/j.ijmedinf.2006.11.006.
27. Meisel, S., and Mattfeld, D., Synergies of operations research and data mining. *Eur. J. Oper. Res.* 206(1):1–10, 2010. doi:10.1016/j.ejor.2009.10.017.
28. Ramesh, A. N., Kambhampati, C., Monson, J. R., and Drew, P. J., Artificial intelligence in medicine. *Ann. R. Coll. Surg. Engl.* 86(5):334–338, 2004. doi:10.1308/147870804290.
29. Ting, H., Mai, Y. T., Hsu, H. C., Wu, H. C., and Tseng, M. H., Decision tree based diagnostic system for moderate to severe obstructive sleep apnea. *J. Med. Syst.* 38(9):94, 2014. doi:10.1007/s10916-014-0094-1.
30. Keltch, B., Lin, Y., and Bayrak, C., Comparison of AI techniques for prediction of liver fibrosis in hepatitis patients. *J. Med. Syst.* 38(8):60, 2014. doi:10.1007/s10916-014-0060-y.
31. Sen, B., Peker, M., Cavusoglu, A., and Celebi, F. V., A comparative study on classification of sleep stage based on EEG signals using feature selection and classification algorithms. *J. Med. Syst.* 38(3):18, 2014. doi:10.1007/s10916-014-0018-0.
32. Maas, A. I., Hukkelhoven, C. W., Marshall, L. F., and Steyerberg, E. W., Prediction of outcome in traumatic brain injury with computed tomographic characteristics: A comparison between the computed tomographic classification and combinations of computed tomographic predictors. *Neurosurgery* 57(6):1173–1182, 2005. discussion 1173–1182.
33. Segal, M. E., Goodman, P. H., Goldstein, R., Hauck, W., Whyte, J., Graham, J. W., Polansky, M., and Hammond, F. M., The accuracy of artificial neural networks in predicting long-term outcome after traumatic brain injury. *J. Head Trauma Rehabil.* 21(4):298–314, 2006.
34. Lee, S. Y., Kim, S. S., Kim, C. H., Park, S. W., Park, J. H., and Yeo, M., Prediction of outcome after traumatic brain injury using clinical and neuroimaging variables. *J. Clin. Neurol.* 8(3):224–229, 2012. doi:10.3988/jcn.2012.8.3.224.
35. Jennett, B., Teasdale, G., Galbraith, S., Pickard, J., Grant, H., Braakman, R., Avezaat, C., Maas, A., Minderhoud, J., Vecht, C. J., Heiden, J., Small, R., Caton, W., and Kurze, T., Severe head injuries in three countries. *J. Neurol. Neurosurg. Psychiatry* 40(3):291–298, 1977.
36. Jennett, B., Snoek, J., Bond, M. R., and Brooks, N., Disability after severe head injury: Observations on the use of the Glasgow Outcome Scale. *J. Neurol. Neurosurg. Psychiatry* 44(4):285–293, 1981.
37. Overgaard, J., Hvid-Hansen, O., Land, A. M., Pedersen, K. K., Christensen, S., Haase, J., Hein, O., and Tweed, W. A., Prognosis after head injury based on early clinical examination. *Lancet* 2(7830):631–635, 1973.
38. Becker, D. P., Miller, J. D., Ward, J. D., Greenberg, R. P., Young, H. F., and Sakalas, R., The outcome from severe head injury with early

diagnosis and intensive management. *J. Neurosurg.* 47(4):491–502, 1977. doi:10.3171/jns.1977.47.4.0491.
39. Carlsson, C. A., von Essen, C., and Lofgren, J., Factors affecting the clinical corse of patients with severe head injuries. 1. Influence of biological factors. 2. Significance of posttraumatic coma. *J. Neurosurg.* 29(3):242–251, 1968. doi:10.3171/jns.1968.29.3.0242.
40. Bowers, S. A., and Marshall, L. F., Outcome in 200 consecutive cases of severe head injury treated in San Diego County: A prospective analysis. *Neurosurgery* 6(3):237–242, 1980.
41. Levati, A., Farina, M. L., Vecchi, G., Rossanda, M., and Marrubini, M. B., Prognosis of severe head injuries. *J. Neurosurg.* 57(6):779–783, 1982. doi:10.3171/jns.1982.57.6.0779.
42. Mazumdar, M., and Glassman, J. R., Categorizing a prognostic variable: Review of methods, code for easy implementation and applications to decision-making about cancer treatments. *Stat. Med.* 19(1):113–132, 2000.
43. Shrout, P. E., and Fleiss, J. L., Intraclass correlations: Uses in assessing rater reliability. *Psychol. Bull.* 86(2):420–428, 1979.
44. Lin, C. C., Ou, Y. K., Chen, S. H., Liu, Y. C., and Lin, J., Comparison of artificial neural network and logistic regression models for predicting mortality in elderly patients with hip fracture. *Injury* 41(8):869–873, 2010. doi:10.1016/j.injury.2010.04.023.
45. Patterson, D. W., *Artificial neural networks: theory and applications*. Prentice Hall, Singapore, New York, 1996.
46. Quinlan, J. R., *C4.5: programs for machine learning*. Morgan Kaufmann, San Francisco, 1993.
47. Breiman, L., *Classification and regression trees. Wadsworth statistics/probability series*. Wadsworth International Group, Belmont, Calif., 1984.
48. Podgorelec, V., Kokol, P., Stiglic, B., and Rozman, I., Decision trees: An overview and their use in medicine. *J. Med. Syst., Kluwer Academic/Plenum Press* 26(5):445–463, 2002.
49. Quinlan, J. R., Induction of decision trees. *Mach. Learn.* 1:81–106, 1986.
50. Polat, K., Sahan, S., and Gunes, S., A new method to medical diagnosis: Artificial immune recognition system (AIRS) with fuzzy weighted pre-processing and application to ECG arrhythmia. *Expert Syst. Appl.* 31(2):264–269, 2006. doi:10.1016/j.eswa.2005.09.019.
51. Bengio, Y., and Grandvalet, Y., No unbiased estimator of the variance of K-fold cross-validation. *J. Mach. Learn. Res.* 5:1089–1105, 2004.
52. Stone, M., Cross-validatory choice and assessment of statistical predictions. *J. R. Stat. Soc.* 36(1):111–147, 1974.
53. Efron, B., and Tibshirani, R., Improvements on cross-validation: The. 632+ bootstrap method. *J. Am. Stat. Assoc.* 92(438):548–560, 1997.
54. Frawley, W. J., Paitetsky-Shapiro, G., and Matheus, C. J. Advances in knowledge discovery and data mining. In: Fayyad, U. M. (Ed.), *From Data Mining to Knowledge Discovery: An Overview.* AAAI Press/ The MIT Press, p 661–620, 1996.
55. Breiman, L., Friedman, J., Olshen, R., and Stone, C., *Classification and regression trees*. Wadsworth, Belmont, CA, 1984.
56. Dreiseitl, S., and Ohno-Machado, L., Logistic regression and artificial neural network classification models: A methodology review. *J. Biomed. Inform.* 35(5–6):352–359, 2002.
57. Altman, D. G., and Bland, J. M., Diagnostic tests 3: Receiver operating characteristic plots. *BMJ* 309(6948):188, 1994.
58. Akobeng, A. K., Understanding diagnostic tests 1: Sensitivity, specificity and predictive values. *Acta Paediatr.* 96(3):338–341, 2007. doi:10.1111/j.1651-2227.2006.00180.x.
59. Code, C., Rowley, D., and Kertesz, A., Predicting recovery from aphasia with connectionist networks: Preliminary comparisons with multiple regression. *Cortex* 30(3):527–532, 1994.
60. Narayan, R. K., Greenberg, R. P., Miller, J. D., Enas, G. G., Choi, S. C., Kishore, P. R., Selhorst, J. B., Lutz, H. A., 3rd, and Becker, D. P., Improved confidence of outcome prediction in severe head injury. A comparative analysis of the clinical examination, multimodality evoked potentials, CT scanning, and intracranial pressure. *J. Neurosurg.* 54(6):751–762, 1981. doi:10.3171/jns.1981.54.6.0751.
61. Hoffmann, M., Lefering, R., Rueger, J. M., Kolb, J. P., Izbicki, J. R., Ruecker, A. H., Rupprecht, M., Lehmann, W., and Trauma Registry of the German Society for Trauma S, Pupil evaluation in addition to Glasgow Coma Scale components in prediction of traumatic brain injury and mortality. *Br. J. Surg.* 99(Suppl 1):122–130, 2012. doi:10.1002/bjs.7707.
62. Signorini, D. F., Andrews, P. J., Jones, P. A., Wardlaw, J. M., and Miller, J. D., Predicting survival using simple clinical variables: A case study in traumatic brain injury. *J. Neurol. Neurosurg. Psychiatry* 66(1):20–25, 1999.
63. Lesko, M. M., Jenks, T., O'Brien, S. J., Childs, C., Bouamra, O., Woodford, M., and Lecky, F., Comparing model performance for survival prediction using total Glasgow Coma Scale and its components in traumatic brain injury. *J. Neurotrauma* 30(1):17–22, 2013. doi:10.1089/neu.2012.2438.
64. Nelson, D. W., Rudehill, A., MacCallum, R. M., Holst, A., Wanecek, M., Weitzberg, E., and Bellander, B. M., Multivariate outcome prediction in traumatic brain injury with focus on laboratory values. *J. Neurotrauma* 29(17):2613–2624, 2012. doi:10.1089/neu.2012.2468.
65. Leitgeb, J., Mauritz, W., Brazinova, A., Majdan, M., Janciak, I., Wilbacher, I., and Rusnak, M., Glasgow Coma Scale score at intensive care unit discharge predicts the 1-year outcome of patients with severe traumatic brain injury. *Eur. J. Trauma Emerg. Surg.* 39(3):285–292, 2013. doi:10.1007/s00068-013-0269-3.
66. Deeks, J. J., and Altman, D. G., Diagnostic tests 4: Likelihood ratios. *BMJ* 329:168–169, 2004.
67. Dujardin, B., Ende, J. V., Gompel, A. V., Unger, J.-P., and Stuyft, P. V. D., Likelihood ratios: A real improvement for clinical decision making? *Eur. J. Epidemiol.* 10:29–36, 1994.
68. Gill, C. J., Sabin, L., and Schmid, C. H., Why clinicians are natural Bayesians. *BMJ* 330(7499):1080–1083, 2005.
69. Akobeng, A. K., Understanding diagnostic tests 2: Likelihood ratios, pre- and post-test probabilities and their use in clinical practice. *Acta Paediatr.* 96(4):487–491, 2007. doi:10.1111/j.1651-2227.2006.00179.x.