Prediction of outcome after severe and moderate head injury by classification and regression tree (CART) technique

Vineet kumar kamal[1], R. M. Pandey[1,] Deepak Agrawal[2]
[1]*Department of Biostatistics,* [2]*Department of Neurosurgery,*
Neurosciences & Gamma Knife Centre, All India Institute of Medical Sciences, New Delhi - 110029, India.

Abstract:
For developing and validating a prognostic model for in-hospital mortality and unfavourable outcome at 6-months in moderate and severe head injury patients, a CART technique was employed in the analysis of a tertiary care trauma database (n=1466 patients) by using 24 prognostic indicators. For in-hospital mortality, there were 7 terminal nodes and the area under curve w as 0.83 and 0.82 for learning and test data sample respectively. The overall classification predictive accuracy was 82% for learning data sample and 79% for test data sample. For 6-months outcome, there were 4 terminal nodes and the AUC was 0.82 and 0.79 for learning and test data sample respectively. The overall classification predictive accuracy was 79% for learning data sample and 76% for test data sample. Methodologically, CART is quite different from the more commonly used statistical methods with the primary benefit of illustrating the important prognostic variables as related to outcome.
Key Words: Glasgow Outcome Scale; prognostic models; CART; traumatic brain injury; validation; outcome.

Introduction:

Traumatic brain injury (TBI) poses a leading cause of disability and mortality in all regions of the globe despite advancement in prevention and treatments. TBI is a significant public health problem worldwide and it is predicted to surpass many diseases as a major cause of death and disability by the year 2020 (“World Health Organization. Projections of Mortality," 2002). The incidence of TBI in the United States and Europe has been estimated at between 180 and 250 and up to 500 per 100,000 populations per year respectively (Bruns & Hauser, 2003; Maas, Marmarou, Murray, Teasdale, & Steyerberg, 2007). TBI is the main cause of one third to one half of all trauma deaths and the leading cause of disability in people under 40, severely disabling 15-20 per 100,000 populations per year (Fleminger 2005). Every five minutes, someone dies from a head injur, Every 5 minutes someone becomes permanently disabled due to a head injury and the cost of severe brain injury often exceeds 4 million dollars (http://www.headinjuryctr-stl.org/statistics.html). In comparison to all other global region, Asia has the highest percentage of TBI-related outcomes as a result of falls (77%), unintentional Injuries (57%) and road traffic accidents (48%) (Adnan & Puvanachandra, 2009).

TBI is a leading cause of mortality, morbidity, disability, and socioeconomic losses in India as well as in other developing countries. It is estimated that nearly 1.5 to 2 million persons are injured and 1 million die every year in India (Gururaj 2002). India and other developing countries are facing the major challenges of prevention, pre-hospital care and rehabilitation in their rapidly changing environments to reduce the burden of TBIs (2002).

Like diagnosis and treatment, prognosis is a fundamental responsibility of all clinicians after a TBI in keeping view of patients and families. Statistical modelling is essential for many purposes in TBI. These days, Statistical modelling has been used for prognostication, hypothesis generation and stratification of patients in research studies (Helmy, Timofeev, & Hutchinson, 2010). Accurate prognostication can help in justifiable transfer to neurosurgical specialist services as well as in early management of the individual patient and to advice patient's relatives. We can say that the intelligent application of statistical models can improve our understanding of the pathology and treatment of TBI (2010).

Existing literatures show that very few studies have been done and none of the studies have developed prognostic models for prediction of outcome in TBI patients in India. Most of the previously developed models may not be well suited to India and other similar counties because they are based on western setting and population. So, the generalizability and applicability of previously developed models for outcome prediction to these settings are limited.

Many studies have constructed mathematical and statistical predictive models that describe and quantify the relationship between possible prognostic factor and outcome for head injury patients. Most studies have used some form of linear regression, with the results presented as a regression equation. Most of them contained relatively complex formulae requiring computers, or the use of expensive, time consuming, or highly specialised measurements. Many of the previous prediction models do not provide information about the critical point-thresholds of each indicators beyond which the risk of a good outcome is substantially increased or decreased.

CART is an alternative statistical method of making predictions from data based on repeated partitioning of the dataset into more homogeneous subgroup(Leo Breiman et al. 1984)(D. Steinberg and P. Colla 1995)(Steinberg, P. Colla, and K. Martin 1997). Results from CART are presented as "decision trees" that require no calculation for their use. Other desirable properties of CART include incorporation of nonlinear relationships and interactions and the ability to predict outcome of cases despite some missing data. Earlier, Choi, *et al*., used CART to predict GOS scores for severely injured patients.

CART is a relatively new, advanced tool for tree-structured data analysis. Although the theory and the mathematical algorithms of the technique are quite complex, the CART program requires no special training to use. CART uses a decision tree to display how data may be classified or predicted. Through a series of yes/no Questions concerning database fields, CART automatically searches for important relationships and uncovers hidden structure even in highly complex data. CART is often used to select a manageable number of core measures from databases with hundreds of variables. Because CART works automatically, even on complex data sets, producing results that are easy to understand, it is being used increasingly in medical, marketing, environmental, banking and commercial applications. In the last ten years, several hundred scholarly articles have referred to the CART methodology (D. Steinberg and P. Colla 1995).

Therefore, the objective of this study was to develop a simple model based on Classification and regression tree (CART) technique for prediction of In-hospital mortality and unfavourable outcome based on Glasgow Outcome Scale (GOS) (Jennett and Bond 1975) at 6-months post trauma after severe and moderate head injury that involves a set of variables that are rapidly and easily achievable in routine neurosurgerical practice using admission characteristics.

Materials and Methods

Patients: We included all the patients with moderate and severe head injury, i.e. the patients having admission Glasgow coma scale (GCS) ≤ 12 in Neurosurgery Casualty Department and admitted to ICU at Jai Prakash Narayan Apex Trauma Center (JPNATC), AIIMS, New Delhi, within first 72 hours of injury during June 19, 2010 to July 31, 2012.

Predictors and Outcome: We considered patients' characteristics which were previously reported as important predictors in literatures and that could be determined easily and reliable within the first hours after injury (Lingsma et al. 2010). These included the information based on demographics (age, sex), clinical severity (cause of injury, the motor GCS at admission, Pupil reactivity and limb movement), secondary insult (hypotension), various CT findings (midline shift, SDH, EDH, basal cistern effaced, presence of tSAH), various blood results (hemoglobin, glucose level, sodium, creatinine).

Statistical analysis: For collecting data, a Performa was prepared in software EPI Info 7.1.2 for all the patients. Prior to analysis, extensive data checking and data cleaning were performed to determine incompleteness, incorrectness, inaccuracy, entry of the data and removing errors from the data before doing analysis. It is well known that a certain combination of factors yields a more effective prediction of outcome instead of using a single factor. A total of 24 prognostic indicators were examined to predict In-hospital mortality and outcome at 6 months after head injury.

The cart technique is an alternative method over other tradition method of prediction (Leo Breiman et al. 1984)(D. Steinberg and P. Colla 1995)(Steinberg, P. Colla, and K. Martin 1997). It is based on statistically optimum recursive splitting of the patients into smaller and smaller sub-groups using some critical level of the prognostic variables. In this method, the dataset is split into two subgroups that are the most different with respect to the outcome. This process is continued for each subgroup until some minimum subgroup size is reached.

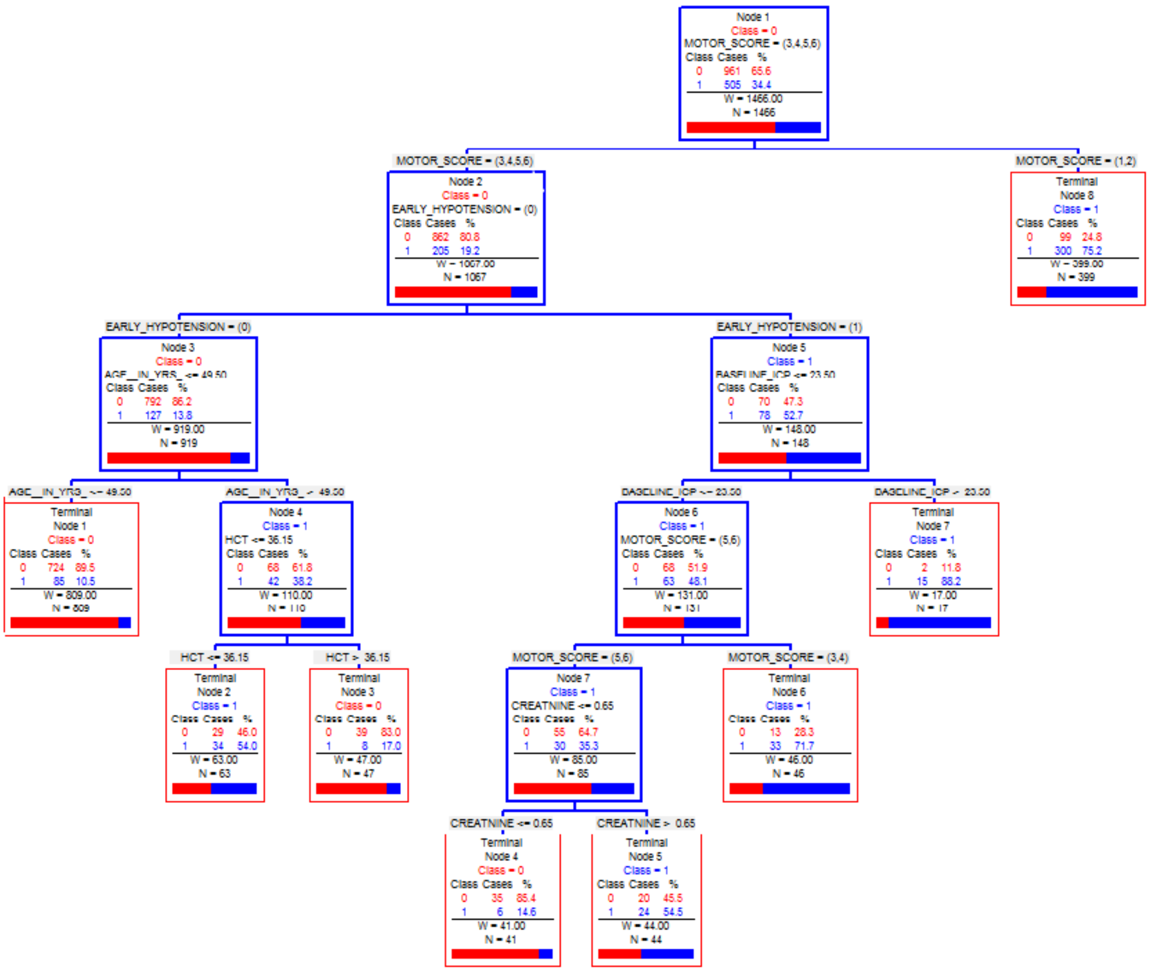
The selected splitting method for growing the classification tree was Gini method in this analysis with other condition of having at least 10 patients at each of the final subgroups. Cross-validation method was used to assess the performance of the prediction tree and the independent predictive accuracy of the model. The tree, which minimizes the overall cross-validated relative error estimate, was presented which most accurately predicts data excluded from forming the tree.

Results: A total of 24 prognostic indicators were examined to predict In-hospital mortality and outcome at 6 months after head injury. For In-hospital mortality, there were 7 terminal nodes and the area under curve w as 0.83 and 0.82 for learning and test data sample respectively. The overall classification predictive accuracy was 82% for learning data sample and 79% for test data sample, with a relative cost 0.37 for learning data sample. For 6-months outcome, there w ere 4 terminal nodes and the area under curve w as 0.82 and 0.79 for learning and test data sample respectively. The overall classification predictive accuracy was 79% for learning data sample and 76% for test data sample, with a relative cost 0.40 for learning data sample

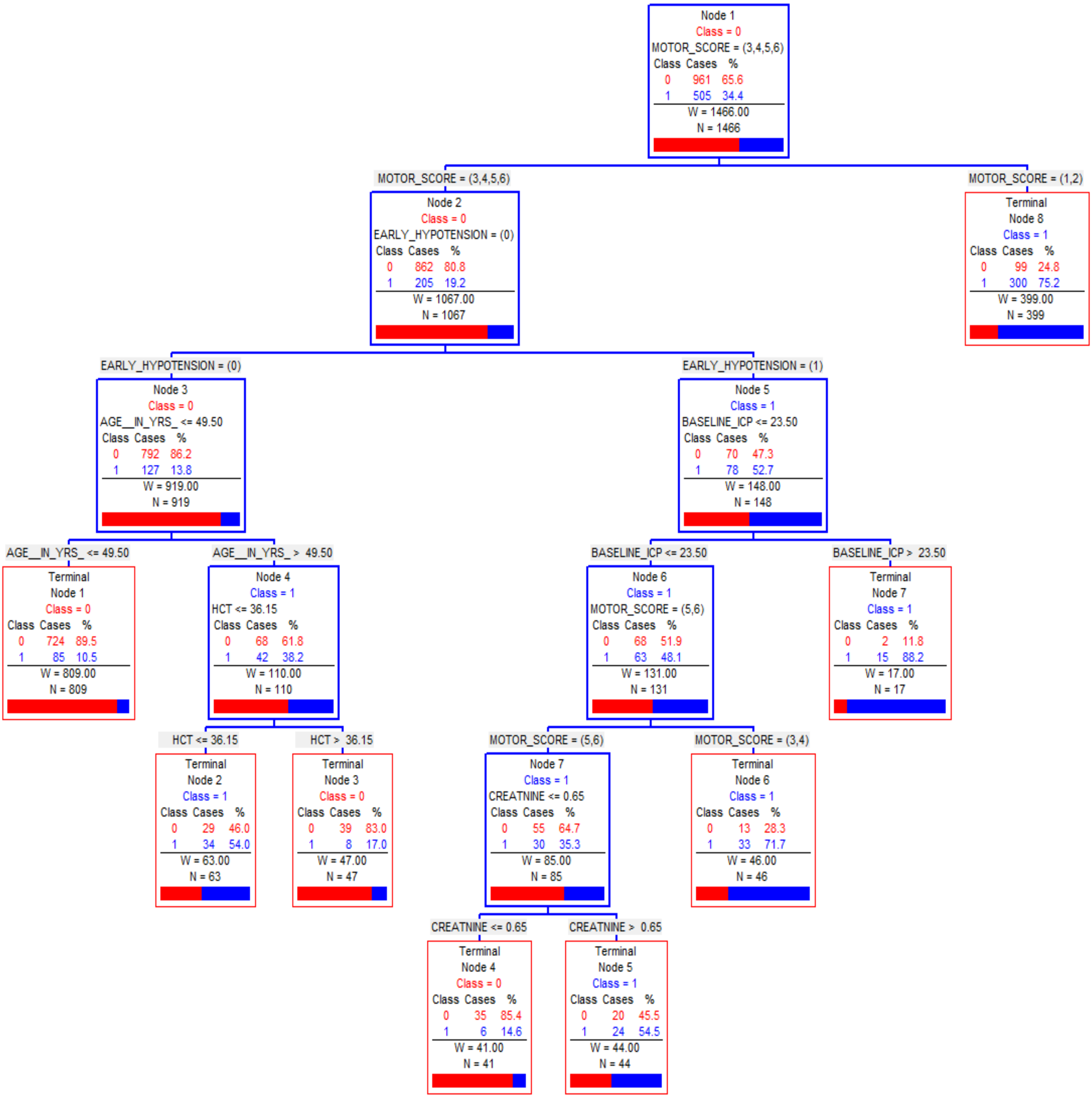
Conclusions: Methodologically, CART is quite different from the more commonly used statistical methods with the primary benefit of illustrating the important prognostic variables as related to outcome. This is very easy for clinical understanding. This seems less expensive, less time consuming, and less specialized measurements and may prove useful in developing new therapeutic strategies and approaches. We are the first in India to develop the CART model for head injury patients based on the largest sample size ever in India.

Tables and Figures:

For In-hospital mortality:



For unfavourable outcome:



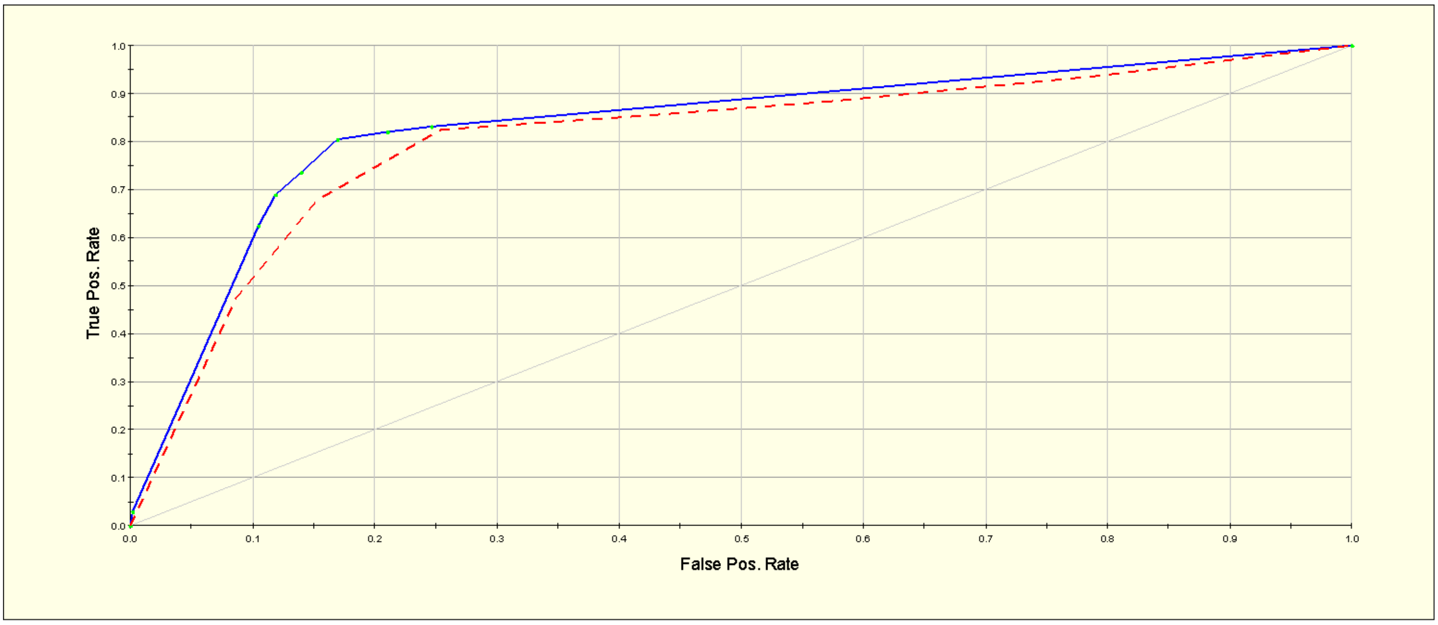
For In-hospital mortality:

Summary

Name	Learn	Test
Average LogLikelihood (Negative)	0.43743	0.46615
Misclass Rate Overall (Raw)	0.17872	0.20873
ROC (Area Under Curve)	0.83365	0.82142
Lift	2.22661	2.16601
Class. Accuracy (Baseline threshold)	0.82128	0.78718

Relative Cost	0.36565	0.42614

CART Navigator 10 (8 Nodes) - Summary Results - Gains Chart - ROC, Sample: Full sample, Target class: 1



Variable Importance

Variable	Score	
EARLY_HYPOTENSION	100.0000	||||||||||||||||||||||||||||||||||||||||
MOTOR_SCORE	83.4997	|||||||||||||||||||||||||||||||||
PUPIL_REACTIVITY	59.4103	|||||||||||||||||||||||
LIMB_MOVEMENT	30.8246	||||||||||||
CREATNINE	16.1538	||||||
AGE__IN_YRS_	13.8861	|||||
RBC	6.5078	|||
SODIUM	6.3322	|||
BASELINE_ICP	5.7059	||
HB	5.1795	||
HCT	3.7710		
UREA_NITROGEN	3.0191		
PLATELET	1.1757	

TLC	0.5450	
CAUSE_OF_INJURY	0.1682	
GLUCOSE	0.1522	

For Unfavorable Outcome:

Summary

Name	Learn	Test
Average LogLikelihood (Negative)	0.43743	0.46615
Misclass Rate Overall (Raw)	0.17872	0.20873
ROC (Area Under Curve)	0.83365	0.82142
Lift	2.22661	2.16601
Class. Accuracy (Baseline threshold)	0.82128	0.78718
Relative Cost	0.36565	0.42614

Variable Importance

Variable	Score	
EARLY_HYPOTENSION	100.0000	
MOTOR_SCORE	83.4997	
PUPIL_REACTIVITY	59.4103	
LIMB_MOVEMENT	30.8246	
CREATNINE	16.1538	
AGE__IN_YRS_	13.8861	
RBC	6.5078	
SODIUM	6.3322	

BASELINE_ICP	5.7059	
HB	5.1795	
HCT	3.7710	
UREA_NITROGEN	3.0191	
PLATELET	1.1757	
TLC	0.5450	
CAUSE_OF_INJURY	0.1682	
GLUCOSE	0.1522	

References:

Adnan A. Hyder and Prasanthi Puvanachandra. 2009. “THE BURDEN OF TRAUMATIC BRAIN INJURY IN ASIA: A CALL FOR RESEARCH.” *Pakistan Jounal of Neurological Sciences* 4(1) (March): 7–32.

Brain Trauma Foundation, American Association of Neurological Surgeons, and Congress of Neurological Surgeons. 2007. “Guidelines for the Management of Severe Traumatic Brain Injury.” *Journal of Neurotrauma* 24 Suppl 1: S1–106. doi:10.1089/neu.2007.9999.

Bruns, John, Jr, and W Allen Hauser. 2003. “The Epidemiology of Traumatic Brain Injury: A Review.” *Epilepsia* 44 Suppl 10: 2–10.

D. Steinberg, and P. Colla. 1995. “Tree-Structured Non-Parametric Data analysis.San Diego, Calif., U.S.A”. Salford Systems.

Diamond, G A. 1992. “What Price Perfection? Calibration and Discrimination of Clinical Prediction Models.” *Journal of Clinical Epidemiology* 45 (1): 85–89.

Fleminger, S. 2005. “Long Term Outcome after Traumatic Brain Injury.” *BMJ* 331 (7530): 1419–20. doi:10.1136/bmj.331.7530.1419.

Gururaj, G. 2002. “Epidemiology of Traumatic Brain Injuries: Indian Scenario.” *Neurological Research* 24 (1): 24–28.

Harrell, F E, Jr, K L Lee, and D B Mark. 1996. “Multivariable Prognostic Models: Issues in Developing Models, Evaluating Assumptions and Adequacy, and Measuring and Reducing Errors.” *Statistics in Medicine* 15 (4): 361–87. doi:10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4.

Harrell FE. 2001. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. New York: springer.

Hukkelhoven, Chantal W P M, Anneke J J Rampen, Andrew I R Maas, Elana Farace, J Dik F Habbema, Anthony Marmarou, Lawrence F Marshall, Gordon D Murray, and Ewout W Steyerberg. 2006. “Some Prognostic Models for Traumatic Brain Injury Were Not Valid.” *Journal of Clinical Epidemiology* 59 (2): 132–43. doi:10.1016/j.jclinepi.2005.06.009.

Jennett, B, and M Bond. 1975. “Assessment of Outcome after Severe Brain Damage.” *Lancet* 1 (7905): 480–84.

Jennett, B, G Teasdale, R Braakman, J Minderhoud, and R Knill-Jones. 1976. “Predicting Outcome in Individual Patients after Severe Head Injury.” *Lancet* 1 (7968): 1031–34.

Junqué, C, O Bruna, and M Mataró. 1997. “Information Needs of the Traumatic Brain Injury Patient’s Family Members Regarding the Consequences of the Injury and Associated Perception of Physical, Cognitive, Emotional and Quality of Life Changes.” *Brain Injury: [BI]* 11 (4): 251–58.

Leo Breiman, Jerome Friedman, Charles J. Stone, and R.A. Olshen. 1984. *Classification and Regression Trees*. The Wadsworth Statistics/Probability Series. Chapman and Hall/CRC.

Lingsma, Hester F, Bob Roozenbeek, Ewout W Steyerberg, Gordon D Murray, and Andrew I R Maas. 2010. “Early Prognosis in Traumatic Brain Injury: From Prophecies to Predictions.” *Lancet Neurology* 9 (5): 543–54. doi:10.1016/S1474-4422(10)70065-X.

Marmarou, Anthony, Juan Lu, Isabella Butcher, Gillian S McHugh, Gordon D Murray, Ewout W Steyerberg, Nino A Mushkudiani, Sung Choi, and Andrew I R Maas. 2007. “Prognostic Value of the Glasgow Coma Scale and Pupil Reactivity in Traumatic Brain Injury Assessed Pre-Hospital and on Enrollment: An IMPACT Analysis.” *Journal of Neurotrauma* 24 (2): 270–80. doi:10.1089/neu.2006.0029.

Moons, Karel G M, Frank E Harrell, and Ewout W Steyerberg. 2002. “Should Scoring Rules Be Based on Odds Ratios or Regression Coefficients?” *Journal of Clinical Epidemiology* 55 (10): 1054–55.

Perel, Pablo, Phil Edwards, Reinhard Wentz, and Ian Roberts. 2006. “Systematic Review of Prognostic Models in Traumatic Brain Injury.” *BMC Medical Informatics and Decision Making* 6: 38. doi:10.1186/1472-6947-6-38.

Schreiber, Martin A, Noriaki Aoki, Bradford G Scott, and J Robert Beck. 2002. “Determinants of Mortality in Patients with Severe Blunt Head Injury.” *Archives of Surgery (Chicago, Ill.: 1960)* 137 (3): 285–90.

Servadei, F, G D Murray, K Penny, G M Teasdale, M Dearden, F Iannotti, F Lapierre, et al. 2000. “The Value of the ‘Worst’ Computed Tomographic Scan in Clinical Studies of Moderate and Severe Head Injury. European Brain Injury Consortium.” *Neurosurgery* 46 (1): 70–75; discussion 75–77.

Steinberg, D., P. Colla, and K. Martin. 1997. “CART—Classification and Regression Trees: Supplementary Manual for Windows.San Diego, Calif., U.S.A.” Salford Systems.

Steyerberg, Ewout W, Nino Mushkudiani, Pablo Perel, Isabella Butcher, Juan Lu, Gillian S McHugh, Gordon D Murray, et al. 2008. “Predicting Outcome after Traumatic Brain Injury: Development and International Validation of Prognostic Scores Based on Admission Characteristics.” *PLoS Medicine* 5 (8): e165; discussion e165. doi:10.1371/journal.pmed.0050165.

Van Beek, Jackelien G M, Nino A Mushkudiani, Ewout W Steyerberg, Isabella Butcher, Gillian S McHugh, Juan Lu, Anthony Marmarou, Gordon D Murray, and Andrew I R Maas. 2007. “Prognostic Value of Admission Laboratory Parameters in Traumatic Brain Injury: Results from the IMPACT Study.” *Journal of Neurotrauma* 24 (2): 315–28. doi:10.1089/neu.2006.0034.

“What Is the Purpose of Statistical Modelling in Traumatic Brain Injury?” 2010. *Acta Neurochirurgica* 152 (11): 2007–8.

“World Health Organization. Projections of Mortality and Burden of Disease to 2030: Deaths by Income Group. Geneva: 2002. 12/01/06.”